GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 18, 2004, 21:02:52 ; Search time 49 Seconds
(without alignments)
8607.685 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 1365
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 2327084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10054536/runat_18062004_173457_13682/app_query.fasta_1.903
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10054536_@CGN_1_1_13_@runat_18062004_173457_13682
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1321	96.8	248	14	US-10-054-536-28	Sequence 28, Appl
2	1321	96.8	248	15	US-10-388-322-1	Sequence 1, Appli
3	1321	96.8	248	16	US-10-429-160-58	Sequence 58, Appl
4	1228	90.0	228	12	US-09-971-475-1	Sequence 1, Appli
5	454	33.3	375	12	US-10-312-829-9	Sequence 9, Appli
6	442.5	32.4	351	13	US-10-007-408-1	Sequence 1, Appli
7	441	32.3	374	12	US-10-312-829-7	Sequence 7, Appli
8	404	29.6	248	12	US-10-312-829-12	Sequence 12, Appl
9	398.5	29.2	248	12	US-10-312-829-4	Sequence 4, Appli
10	396.5	29.0	243	12	US-10-336-603A-96	Sequence 96, Appl
11	394.5	28.9	248	12	US-10-336-603A-100	Sequence 100, App
12	389.5	28.5	259	9	US-09-925-302-473	Sequence 473, App
13	389.5	28.5	259	12	US-09-925-302-473	Sequence 473, App
14	314.5	23.0	153	12	US-10-312-829-10	Sequence 10, Appl
15	304.5	22.3	271	9	US-09-989-722-357	Sequence 357, App
16	304.5	22.3	271	9	US-09-989-723-357	Sequence 357, App
17	304.5	22.3	271	9	US-09-989-279-357	Sequence 357, App
18	304.5	22.3	271	9	US-09-989-727-357	Sequence 357, App
19	304.5	22.3	271	9	US-09-989-731-357	Sequence 357, App
20	304.5	22.3	271	9	US-09-989-732-357	Sequence 357, App
21	304.5	22.3	271	9	US-09-991-073-357	Sequence 357, App
22	304.5	22.3	271	9	US-09-990-442-357	Sequence 357, App
23	304.5	22.3	271	9	US-09-991-163-357	Sequence 357, App
24	304.5	22.3	271	9	US-09-993-604-357	Sequence 357, App
25	304.5	22.3	271	9	US-09-990-456-357	Sequence 357, App
26	304.5	22.3	271	9	US-09-989-721-357	Sequence 357, App
27	304.5	22.3	271	9	US-09-992-598-357	Sequence 357, App
28	304.5	22.3	271	9	US-09-989-293A-357	Sequence 357, App
29	304.5	22.3	271	9	US-09-989-735-357	Sequence 357, App
30	304.5	22.3	271	9	US-09-990-444-357	Sequence 357, App
31	304.5	22.3	271	9	US-09-991-181-357	Sequence 357, App
32	304.5	22.3	271	9	US-09-989-730-357	Sequence 357, App
33	304.5	22.3	271	9	US-09-990-436-357	Sequence 357, App
34	304.5	22.3	271	9	US-09-993-687-357	Sequence 357, App
35	304.5	22.3	271	10	US-09-989-734-357	Sequence 357, App
36	304.5	22.3	271	10	US-09-997-653-357	Sequence 357, App
37	304.5	22.3	271	10	US-09-993-667-357	Sequence 357, App
38	304.5	22.3	271	10	US-09-997-428-357	Sequence 357, App
39	304.5	22.3	271	10	US-09-997-666-357	Sequence 357, App
40	304.5	22.3	271	10	US-09-990-438-357	Sequence 357, App
41	304.5	22.3	271	10	US-09-990-562-357	Sequence 357, App
42	304.5	22.3	271	10	US-09-990-711-357	Sequence 357, App
43	304.5	22.3	271	10	US-09-989-726-357	Sequence 357, App
44	304.5	22.3	271	10	US-09-998-156-357	Sequence 357, App
45	304.5	22.3	271	10	US-09-990-437-357	Sequence 357, App

ALIGNMENTS

RESULT 1
US-10-054-536-28
; Sequence 28, Application US/10054536
; Publication No. US20030162248A1
; GENERAL INFORMATION:
; APPLICANT: Wakamiya, No. US20030162248A1utaka
; TITLE OF INVENTION: RECOMBINANT HUMAN MANNAN-BINDING PROTEINS AND PROCESS
; TITLE OF INVENTION: FOR PRODUCING THE SAME
; FILE REFERENCE: 19036/36614
; CURRENT APPLICATION NUMBER: US/10/054,536
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03311
; PRIOR FILING DATE: 1998-07-23
; PRIOR APPLICATION NUMBER: JP 10-11864
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-536-28

Alignment Scores:

Pred. No.:	1.03e-110	Length:	248
Score:	1321.00	Matches:	248
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.78%	Indels:	0
DB:	14	Gaps:	0

US-10-054-536-2 (1-747) x US-10-054-536-28 (1-248)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248
```

RESULT 2
US-10-388-322-1
; Sequence 1, Application US/10388322
; Publication No. US20040006009A1
; GENERAL INFORMATION:
; APPLICANT: NatImmune
; TITLE OF INVENTION: Pharmaceutical compositions comprising mannose binding lectin
; FILE REFERENCE: P 625 DK00
; CURRENT APPLICATION NUMBER: US/10/388,322
; CURRENT FILING DATE: 2003-03-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-388-322-1

Alignment Scores:

Pred. No.:	1.03e-110	Length:	248
Score:	1321.00	Matches:	248
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.78%	Indels:	0
DB:	15	Gaps:	0

US-10-054-536-2 (1-747) x US-10-388-322-1 (1-248)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248
```

RESULT 3
US-10-429-160-58
; Sequence 58, Application US/10429160
; Publication No. US20040023276A1
; GENERAL INFORMATION:

```
;  APPLICANT: Ward, Teresa R
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Linsley, Peter S
;  APPLICANT:  Lund, Lund
;  TITLE OF INVENTION: LXR Ligand Induced Genes and Proteins
;  FILE REFERENCE: RS0200
;  CURRENT APPLICATION NUMBER: US/10/429,160
;  CURRENT FILING DATE:  2003-05-02
;  PRIOR APPLICATION NUMBER: US 60/377,714
;  PRIOR FILING DATE: 2002-05-03
;  NUMBER OF SEQ ID NOS: 111
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-429-160-58

Alignment Scores:
Pred. No.:                1.03e-110     Length:          248
Score:                    1321.00       Matches:         248
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              96.78%        Indels:          0
DB:                       16            Gaps:            0

US-10-054-536-2 (1-747) x US-10-429-160-58 (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              ||||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248

RESULT 4
US-09-971-475-1
; Sequence 1, Application US/09971475
; Publication No. US20020086817A1
; GENERAL INFORMATION:
;  APPLICANT: Kawasaki, Toshisuke
;  TITLE OF INVENTION: Anticancer agent
;  FILE REFERENCE: ADT 308
;  CURRENT APPLICATION NUMBER: US/09/971,475
;  CURRENT FILING DATE:  2001-10-04
;  PRIOR APPLICATION NUMBER: US 09/468,705
;  PRIOR FILING DATE: 1999-12-21
;  PRIOR APPLICATION NUMBER: PCT/JP98/03697
;  PRIOR FILING DATE: 1998-08-19
;  PRIOR APPLICATION NUMBER: JP 239113/97
;  PRIOR FILING DATE: 1997-08-21
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 228
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-971-475-1

Alignment Scores:
Pred. No.:                2.58e-102     Length:          228
Score:                    1228.00       Matches:         228
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              89.96%        Indels:          0
DB:                       12            Gaps:            0

US-10-054-536-2 (1-747) x US-09-971-475-1 (1-228)

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 20

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 40

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 60

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 80

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 100

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 120

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 140

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 160
```

```
Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 180

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 200

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 220

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        221 LeuAlaValCysGluPheProIle 228
```

RESULT 5
US-10-312-829-9
; Sequence 9, Application US/10312829
; Publication No. US20040037781A1
; GENERAL INFORMATION:
; APPLICANT: McCormack, Francis X
; TITLE OF INVENTION: Peptides with Antioxidant and Antimicrobial Properties
; FILE REFERENCE: 10738-31
; CURRENT APPLICATION NUMBER: US/10/312,829
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: PCT/US01/21226
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,313
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-312-829-9

Alignment Scores:

Pred. No.:	2.83e-32	Length:	375
Score:	454.00	Matches:	99
Percent Similarity:	55.07%	Conservative:	26
Best Local Similarity:	43.61%	Mismatches:	82
Query Match:	33.26%	Indels:	20
DB:	12	Gaps:	5

US-10-054-536-2 (1-747) x US-10-312-829-9 (1-375)

```
Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||:::   |||      |||      |||     |||  |||||||||:::
Db        149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||      |||      |||     |||
Db        169 GlyValProGlyGluArgGlyValProGlyAsnThrGlyAlaAlaGlySerAlaGlyAla 188

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||:::||||||      |||   ||||||  ||||||  ||||||
Db        189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy        286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
              ||||||   |||         |||:::|||   |||         ||||||  |||
Db        209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy        337 AAAGCT----------------------------CTGCAAACAGAAATGGCACGTATC 366
              :::                              ||||||      ::::::
Db        229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||       |||     |||:::  ||||||  |||  |||  |||
Db        249 LysLysValGluLeuProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::      |||   :::  :::  ||||||  :::           :::
Db        269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
              |||:::||||||:::|||||||||||||   |||:::|||   |||:::
Db        289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              |||   |||||||||   :::||||||   ||||||||||:::|||         |||
Db        309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
              |||       |||   |||:::||||||      ||||||||||:::   |||
Db        329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySer 348

Qy        658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              ||||||||||||   :::      ||||||:::|||||||||      |||
Db        349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy        718 CATCTGGCCGTCTGTGAGTTC 738
                  |||   ||||||||||||
Db        369 ArgLeuValValCysGluPhe 375
```

RESULT 6
US-10-007-408-1
; Sequence 1, Application US/10007408
; Publication No. US20020168627A1
; GENERAL INFORMATION:
; APPLICANT: WAKAMIYA, No. US20020168627A1utaka
; TITLE OF INVENTION: RECOMBINANT CONGLUTININ AND PRODUCING
; METHOD THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/007,408
; FILING DATE: 08-No. US20020168627A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP95/02035
; FILING DATE: 02-OCT-1995
; APPLICATION NUMBER: JPA - 209698
; FILING DATE: 17-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 19036/34546
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-007-408-1

Alignment Scores:

Pred. No.:	3.05e-31	Length:	351
Score:	442.50	Matches:	91

```
Percent Similarity:     56.77%        Conservative:  39
Best Local Similarity:  39.74%        Mismatches:    74
Query Match:            32.42%        Indels:        25
DB:                     13            Gaps:          6

US-10-054-536-2 (1-747) x US-10-007-408-1 (1-351)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::|||||||||   ||||||||||||                |||   ||||||||||||
Db        126 AlaProGlyIleGlnGlyPheProGlyPro---------SerGlyLeuLysGlyGluLys 142

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||     |||:::   ||||||   |||
Db        143 GlyAlaProGlyGluThrGlyAlaProGlyArgAlaGlyValThrGlyProSerGlyAla 162

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||       |||       |||   ||||||   ||||||   :::|||
Db        163 IleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGly 182

Qy        286 GACCCTGGAAAAAGT---CCGGATGGTGATAGTAGCCTGGCTGCC-------------- 327
              |||||||||::::::         |||:::|||   ||||||
Db        183 AspProGlyGluThrGlyAlaSerGlyGluSerGlyLeuAlaGluValAsnAlaLeuLys 202

Qy        328 --------------------TCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
                                  :::      |||         ::::::
Db        203 GlnArgValThrIleLeuAspGlyHisLeuArgArgPheGlnAsnAlaPheSerGlnTyr 222

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||   :::   |||      |||:::   ||||||   |||   |||   |||
Db        223 LysLysAlaValLeuPheProAspGlyGlnAlaValGlyGluLysIlePheLysThrAla 242

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::   ::::::         :::   ||||||   :::   :::       :::
Db        243 GlyAlaValLysSerTyrSerAspAlaGluGlnLeuCysArgGluAlaLysGlyGlnLeu 262

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG--------- 537
              |||:::||||||::::::||||||||   |||:::      ::::::::
Db        263 AlaSerProArgSerSerAlaGluAsnGluAlaValThrGlnMetValArgAlaGlnGlu 282

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              :::   |||:::|||   :::   |||       |||||||||:::|||          |||
Db        283 LysAsnAlaTyrLeuSerMetAsnAspIleSerThrGluGlyArgPheThrTyrProThr 302

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT------GGT 651
              |||       |||   |||:::||||||   :::|||||||||||||||:::      |||
Db        303 GlyGluIleLeuValTyrSerAsnTrpAlaAspGlyGluProAsnAsnSerAspGluGly 322

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
                  |||:::||||||   :::      :::|||:::|||||||||||||||||||||
Db        323 GlnProGluAsnCysValGluIlePheProAspGlyLysTrpAsnAspValProCysSer 342

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
                     |||   :::||||||||||
Db        343 LysGlnLeuLeuValIleCysGluPhe 351

RESULT 7
US-10-312-829-7
; Sequence 7, Application US/10312829
; Publication No. US20040037781A1
; GENERAL INFORMATION:
;  APPLICANT: McCormack, Francis X
;  TITLE OF INVENTION: Peptides with Antioxidant and Antimicrobial Properties
;  FILE REFERENCE: 10738-31
;  CURRENT APPLICATION NUMBER: US/10/312,829
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: PCT/US01/21226
;  PRIOR FILING DATE: 2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215,313
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Rat
US-10-312-829-7

Alignment Scores:
Pred. No.:              4.24e-31      Length:        374
Score:                  441.00        Matches:       110
Percent Similarity:     38.81%        Conservative:  34
Best Local Similarity:  29.65%        Mismatches:    99
Query Match:            32.31%        Indels:        128
DB:                     12            Gaps:          7

US-10-054-536-2 (1-747) x US-10-312-829-7 (1-374)

Qy         10 TTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTG 69
              |||   |||:::   :::||||||:::      :::             :::|||
Db          4 PheLeuSerMetLeuValLeuLeuValGlnProLeuGlyAspLeuGlyAlaGluMetLys 23

Qy         70 ACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATC 129
              |||                  |||           :::          ||||||
Db         24 ThrLeuSerGlnArgSerIleThrAsnThrCysThrLeuValLeuCysSerProThrGlu 43

Qy        130 AACGGCTTCCCAGGCAAAGATGGGCGTGAT---------GGCACCAAGGGAGAAAAGGGG 180
              ||||||   ||||||:::||||||||||||         |||   :::||||||||||||
Db         44 AsnGlyLeuProGlyArgAspGlyArgAspGlyArgGluGlyProArgGlyGluLysGly 63

Qy        181 GAACCAGGC---------------------------------------------------- 189
              :::||||||
Db         64 AspProGlyLeuProGlyProMetGlyLeuSerGlyLeuProGlyProArgGlyProVal 83

Qy        190 ----------------------------------------CAAGGGCTCAGAGGCTTA 207
                                                     :::|||   ||||||||
Db         84 GlyProLysGlyGluAsnGlySerAlaGlyGluProGlyProLysGlyGluArgGlyLeu 103

Qy        208 CAGGGCCCCCCT---------------------------------------------- 219
                 |||||||||
Db        104 ValGlyProProGlySerProGlyIleSerGlyProAlaGlyLysGluGlyProSerGly 123

Qy        220 ------GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAG 273
                    |||   :::||||||   |||   |||||||||   |||      |||||||||
Db        124 LysGlnGlyAsnIleGlyProGlnGlyLysProGlyProLysGlyGluAlaGlyProLys 143

Qy        274 GGCCAAAAAGGAGACCCTGGA--------------------AAAAGTCCGGATGGT--- 309
              |||:::   |||   ||||||                   |||   |||   |||
Db        144 GlyGluValGlyAlaProGlyMetGlnGlySerAlaGlyAlaLysGlyProAlaGlyPro 163

Qy        309 ------------------------------------------------------------ 309

Db        164 LysGlyGluArgGlyAlaProGlyGluGlnGlyAlaProGlyAsnAlaGlyAlaAlaGly 183

Qy        309 ------------------------------------------------------------ 309

Db        184 ProAlaGlyProAlaGlyProGlnGlyAlaProGlySerArgGlyProProGlyLeuLys 203

Qy        310 ----------------------------------------------------------GATAGT 315
                                                                         ||||||
Db        204 GlyAspArgGlyAlaProGlyAspArgGlyIleLysGlyGluSerGlyLeuProAspSer 223

Qy        316 AGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC--------- 366
              :::             :::   :::||||||      ::::::   |||:::
Db        224 AlaAlaLeuArgGlnGlnMetGluAlaLeuAsnGlyLysLeuGlnArgLeuGluAlaAla 243

Qy        367 ------------AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
                          ||||||         |||      |||:::   ||||||:::|||
Db        244 PheSerArgTyrLysLysAlaAlaLeuPheProAspGlyGlnSerValGlyAspLysIle 263

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              |||                        ||||||      |||   :::|||   :::
Db        264 PheArgAlaAlaAsnSerGluGluProPheGluAspAlaLysGluMetCysArgGlnAla 283

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
                  :::|||:::||||||:::|||   ||||||   |||:::|||   |||:::
Db        284 GlyGlyGlnLeuAlaSerProArgSerAlaThrGluAsnAlaAlaValGlnGlnLeuVal 303
```

```
Qy        535 ---------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTT 585
Db        304 ThrAlaHisSerLysAlaAlaPheLeuSerMetThrAspValGlyThrGluGlyLysPhe 323

Qy        586 GTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAAT 645
Db        324 ThrTyrProThrGlyGluAlaLeuValTyrSerAsnTrpAlaProGlyGluProAsnAsn 343

Qy        646 GCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCC 705
Db        344 AsnGlyGlyAlaGluAsnCysValGluIlePheThrAsnGlyGlnTrpAsnAspLysAla 363

Qy        706 TGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
Db        364 CysGlyGluGlnArgLeuValIleCysGluPhe 374

RESULT 8
US-10-312-829-12
; Sequence 12, Application US/10312829
; Publication No. US20040037781A1
; GENERAL INFORMATION:
;  APPLICANT: McCormack, Francis X
;  TITLE OF INVENTION: Peptides with Antioxidant and Antimicrobial Properties
;  FILE REFERENCE: 10738-31
;  CURRENT APPLICATION NUMBER: US/10/312,829
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: PCT/US01/21226
;  PRIOR FILING DATE: 2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215,313
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: antioxidant peptide
US-10-312-829-12

Alignment Scores:
Pred. No.:                8.55e-28      Length:          248
Score:                    404.00        Matches:         99
Percent Similarity:       51.34%        Conservative:    35
Best Local Similarity:    37.93%        Mismatches:      99
Query Match:              29.60%        Indels:          28
DB:                       12            Gaps:            9

US-10-054-536-2 (1-747) x US-10-312-829-12 (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
Db          1 MetSerLeuCys---SerLeuAlaPheThrLeuPheLeuThrValValAlaGlyIleLys 19

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
Db         20 CysAsnValThr------AspValCysAlaGlySerProGlyIle------ProGlyAla 35

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
Db         36 ProGlyAsnHisGlyLeuProGlyArgAspGlyArgAspGlyValLysGlyAspProGly 55

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTG------------ 228
Db         56 ProProGlyPro--------------MetGlyProProGlyGlyMetProGlyLeuPro 70

Qy        229 GGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGAC 288
Db         71 GlyArgAspGlyLeuProGlyAlaProGlyAlaProGlyGluArgGlyAspLysGlyGlu 90

Qy        289 CCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAA 348
Db         91 ProGlyGluArg------GlyLeuProGlyPheProAlaTyrLeuAspGluGluLeuGln 108

Qy        349 ACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGC--------------- 393
Db        109 ThrGluLeuTyrGluIleLysHisGlnIleLeuGlnThrMetGlyValLeuSerLeuGln 128

Qy        394 ---------AAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
Db        129 GlySerMetLeuSerValGlyAspLysValPheSerThrAsnGlyGlnSerValAsnPhe 148

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
Db        149 AspThrIleLysGluMetCysThrArgAlaGlyGlyAsnIleAlaValProArgThrPro 168

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
Db        169 GluGluAsnGluAlaIleAlaSerIleAlaLysLysTyrAsnAsnTyrValTyrLeuGly 188

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
Db        189 MetIleGluAspGlnThrProGlyAspPheHisTyrLeuAspGlyAlaSerValAsnTyr 208

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
Db        209 ThrAsnTrpTyrProGlyGluProArgGlyGlnGlyLys---GluLysCysValGluMet 227

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
Db        228 TyrThrAspGlyThrTrpAsnAspArgGlyCysLeuGlnTyrArgLeuAlaValCysGlu 247

Qy        736 TTC 738
Db        248 Phe 248

RESULT 9
US-10-312-829-4
; Sequence 4, Application US/10312829
; Publication No. US20040037781A1
; GENERAL INFORMATION:
;  APPLICANT: McCormack, Francis X
;  TITLE OF INVENTION: Peptides with Antioxidant and Antimicrobial Properties
;  FILE REFERENCE: 10738-31
;  CURRENT APPLICATION NUMBER: US/10/312,829
;  CURRENT FILING DATE:  2003-08-04
;  PRIOR APPLICATION NUMBER: PCT/US01/21226
;  PRIOR FILING DATE: 2001-06-29
;  PRIOR APPLICATION NUMBER: US 60/215,313
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-312-829-4

Alignment Scores:
Pred. No.:                2.69e-27      Length:          248
Score:                    398.50        Matches:         96
Percent Similarity:       50.57%        Conservative:    36
Best Local Similarity:    36.78%        Mismatches:      90
Query Match:              29.19%        Indels:          39
DB:                       12            Gaps:            11

US-10-054-536-2 (1-747) x US-10-312-829-4 (1-248)

Qy         22 CCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGAT 81
Db          5 ProLeuAlaLeuThrLeuIleLeuMetAlaAlaSerGlyAlaAlaCysGluValLysAsp 24

Qy         82 GCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCA 141
Db         25 ValCysValGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuPro 42
```

```
Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
              |||:::|||||||||||||||   ||||||:::   |||   ||||||
Db         43 GlyArgAspGlyArgAspGlyValLysGlyAspProGlyProProGlyPro--------- 59

Qy        202 GGCTTACAGGGCCCCCCTGGAAAGTTGGGG---CCTCCAGGAAAT---------CCAGGG 249
              ||||||||||||:::           ||||||||||||                ||||||
Db         60 ------MetGlyProProGlyGluThrProCysProProGlyAsnAsnGlyLeuProGly 77

Qy        250 CCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGAT 306
              |||   ||||||   :::|||:::||||||:::||||||   :::   |||
Db         78 AlaProGlyValProGlyGluArgGlyGluLysGlyGluProGlyGluArgGlyProPro 97

Qy        307 GGT---------------GATAGTAGCCTGGCTGCCTCA------------GAAAGAAAA 339
              |||               |||      |||   |||:::                  :::
Db         98 GlyLeuProAlaHisLeuAspGluGluLeuGlnAlaThrLeuHisAspPheArgHisGln 117

Qy        340 GCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAA 399
              |||||||||      |||                                |||      |||
Db        118 IleLeuGlnThrArgGlyAla--------------------------LeuSerLeuGln 128

Qy        400 ---------------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                              ||||||   |||   |||   :::||||||:::   :::||||||
Db        129 GlySerIleMetThrValGlyGluLysValPheSerSerAsnGlyGlnSerIleThrPhe 148

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
              :::   ::::::      |||   :::             :::|||   |||||||||
Db        149 AspAlaIleGlnGluAlaCysAlaArgAlaGlyGlyArgIleAlaValProArgAsnPro 168

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
              ||||||   ||||||   :::   :::|||:::                |||:::::|||
Db        169 GluGluAsnGluAlaIleAlaSerPheValLysLysTyrAsnThrTyrAlaTyrValGly 188

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
              :::|||:::      :::   |||   |||            |||      :::   |||
Db        189 LeuThrGluGlyProSerProGlyAspPheArgTyrSerAspGlyThrProValAsnTyr 208

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
              |||||||||      |||||||||          |||      |||   ||||||   :::
Db        209 ThrAsnTrpTyrArgGlyGluProAlaGlyArgGlyLys---GluGlnCysValGluMet 227

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                  :::|||||||||||||||      |||      |||   |||   :::|||||||
Db        228 TyrThrAspGlyGlnTrpAsnAspArgAsnCysLeuTyrSerArgLeuThrIleCysGlu 247

Qy        736 TTC 738
              |||
Db        248 Phe 248
```

RESULT 10
US-10-336-603A-96
; Sequence 96, Application US/10336603A
; Publication No. US20040072997A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-533A
; CURRENT APPLICATION NUMBER: US/10/336,603A
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 96
;   LENGTH: 243
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-336-603A-96

Alignment Scores:
Pred. No.: 4.05e-27 Length: 243

Score:	396.50	Matches: 99
Percent Similarity:	50.38%	Conservative: 34
Best Local Similarity:	37.50%	Mismatches: 86
Query Match:	29.05%	Indels: 45
DB:	12	Gaps: 13

US-10-054-536-2 (1-747) x US-10-336-603A-96 (1-243)

```
Qy         19 CTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAG 78
              |||   |||   |||:::   |||   |||:::                      ||||||
Db          1 LeuAlaLeuThrLeuIleLeuMetAlaAlaSerGly-----------AlaAlaCysGlu 16

Qy         79 ------GATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAAC 132
                   |||                |||   :::              :::||||||   :::
Db         17 ValArgAspValCysValGlySerProGlyIle------ProGlyThrProGlySerHis 34

Qy        133 GGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAA 192
              |||   ||||||:::|||||||||||||||   ||||||:::   |||   ||||||
Db         35 GlyLeuProGlyArgAspGlyArgAspGlyValLysGlyAspProGlyProProGlyPro 54

Qy        193 GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGG---CCTCCAGGAAAT------ 243
                              ||||||||||||:::           ||||||||||||
Db         55 --------------MetGlyProProGlyGluThrProCysProProGlyAsnAsnGly 69

Qy        244 ---CCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA---AAA 297
                 ||||||      |||   ||||||   :::|||:::||||||:::||||||   :::
Db         70 LeuProGlyAlaProGlyValProGlyGluArgGlyGluLysGlyGluProGlyGluArg 89

Qy        298 AGTCCGGATGGT--------------GATAGTAGCCTGGCTGCCTCA------------ 330
                 |||   |||               |||      |||   |||:::
Db         90 GlyProProGlyLeuProAlaHisLeuAspGluGluLeuGlnAlaThrLeuHisAspPhe 109

Qy        331 GAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTG 390
                    :::   |||||||||      |||                          |||
Db        110 ArgHisGlnIleLeuGlnThrArgGlyAla--------------------------Leu 120

Qy        391 GGCAAACAA--------------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATA 435
                 |||                |||||   |||   |||   :::||||||:::
Db        121 SerLeuGlnGlySerIleMetThrValGlyGluLysValPheSerSerAsnGlyGlnSer 140

Qy        436 ATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCC 495
              :::|||||:::   ::::::      |||   :::              :::|||   |||
Db        141 IleThrPheAspAlaIleGlnGluAlaCysAlaArgAlaGlyGlyArgIleAlaValPro 160

Qy        496 AGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCC 546
              ||||||      ||||||   ||||||   :::   :::|||:::              |||
Db        161 ArgAsnProGluGluAsnGluAlaIleAlaSerPheValLysLysTyrAsnThrTyrAla 180

Qy        547 TTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGA 606
              ::::::|||:::|||:::      :::   |||   |||              |||
Db        181 TyrValGlyLeuThrGluGlyProSerProGlyAspPheArgTyrSerAspGlyThrPro 200

Qy        607 CTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGT 666
              :::   ||||||||||||      |||||||||          |||      |||   |||
Db        201 ValAsnTyrThrAsnTrpTyrArgGlyGluProAlaGlyArgGlyLys---GluLysCys 219

Qy        667 GTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCC 726
              |||   :::      :::||||||||||||||||      |||      |||   |||
Db        220 ValGluMetTyrThrAspGlyGlnTrpAsnAspArgAsnCysLeuTyrSerArgLeuThr 239

Qy        727 GTCTGTGAGTTC 738
                 :::|||||||||
Db        240 IleCysGluPhe 243
```

RESULT 11
US-10-336-603A-100
; Sequence 100, Application US/10336603A
; Publication No. US20040072997A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOI
; FILE REFERENCE: 21402-533A
; CURRENT APPLICATION NUMBER: US/10/336,603A

```
;  CURRENT FILING DATE:  2003-01-03
;  PRIOR APPLICATION NUMBER: 09/746,491
;  PRIOR FILING DATE: 2000-12-20
;  PRIOR APPLICATION NUMBER: 10/055,569
;  PRIOR FILING DATE: 2001-10-26
;  NUMBER OF SEQ ID NOS: 169
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 100
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-336-603A-100

Alignment Scores:
Pred. No.:                6.18e-27        Length:          248
Score:                    394.50          Matches:         96
Percent Similarity:       50.57%          Conservative:    36
Best Local Similarity:    36.78%          Mismatches:      90
Query Match:              28.90%          Indels:          39
DB:                       12              Gaps:            11

US-10-054-536-2 (1-747) x US-10-336-603A-100 (1-248)

Qy         22 CCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGAT 81
              ||||||   |||      ::::::   |||:::                  :::|||
Db          5 ProLeuAlaLeuAsnLeuIleLeuMetAlaAlaSerGlyAlaValCysGluValLysAsp 24

Qy         82 GCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCA 141
                       |||   :::          :::||||||   :::|||   |||
Db         25 ValCysValGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuPro 42

Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
              |||:::|||||||||||||||   ||||||:::   |||   ||||||
Db         43 GlyArgAspGlyArgAspGlyValLysGlyAspProGlyProProGlyPro--------- 59

Qy        202 GGCTTACAGGGCCCCCCTGGAAAGTTGGGG---CCTCCAGGAAAT---------CCAGGG 249
                     ||||||||||||:::::      ||||||||||||         ||||||
Db         60 ------MetGlyProProGlyGluMetProCysProProGlyAsnAspGlyLeuProGly 77

Qy        250 CCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGAT 306
                 |||   ||||||      |||:::||||||:::||||||   :::   |||
Db         78 AlaProGlyIleProGlyGluCysGlyGluLysGlyGluProGlyGluArgGlyProPro 97

Qy        307 GGT---------------GATAGTAGCCTGGCTGCCTCA-----------GAAAGAAAA 339
              |||               |||      |||   |||:::                 :::
Db         98 GlyLeuProAlaHisLeuAspGluGluLeuGlnAlaThrLeuHisAspPheArgHisGln 117

Qy        340 GCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAA 399
                 |||||||||      |||                     |||      |||
Db        118 IleLeuGlnThrArgGlyAla-------------------------LeuSerLeuGln 128

Qy        400 --------------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                           ||||||   |||   |||   :::||||||:::   :::|||||||
Db        129 GlySerIleMetThrValGlyGluLysValPheSerSerAsnGlyGlnSerIleThrPhe 148

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
              :::   ::::::      |||   :::          :::|||   |||||||||
Db        149 AspAlaIleGlnGluAlaCysAlaArgAlaGlyGlyArgIleAlaValProArgAsnPro 168

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
              ||||||   ||||||   :::   :::|||:::                |||:::::|||
Db        169 GluGluAsnGluAlaIleAlaSerPheValLysLysTyrAsnThrTyrAlaTyrValGly 188

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
              :::|||:::      :::   |||   |||           |||      :::   |||
Db        189 LeuThrGluGlyProSerProGlyAspPheArgTyrSerAspGlyThrProValAsnTyr 208

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
              |||||||||      |||||||||          |||      |||   ||||||   :::
Db        209 ThrAsnTrpTyrArgGlyGluProAlaGlyArgGlyLys---GluGlnCysValGluMet 227

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                 :::||||||||||||||      |||      |||   |||   :::||||||
Db
```

```
Db        228 TyrThrAspGlyGlnTrpAsnAspArgAsnCysLeuTyrSerArgLeuThrIleCysGlu 247

Qy        736 TTC 738
                 |||
Db        248 Phe 248

RESULT 12
US-09-925-302-473
; Sequence 473, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 473
;   LENGTH: 259
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (20)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (61)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (234)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-473

Alignment Scores:
Pred. No.:                1.77e-26        Length:          259
Score:                    389.50          Matches:         95
Percent Similarity:       50.19%          Conservative:    36
Best Local Similarity:    36.40%          Mismatches:      91
Query Match:              28.53%          Indels:          39
DB:                       9               Gaps:            11

US-10-054-536-2 (1-747) x US-09-925-302-473 (1-259)

Qy         22 CCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGAT 81
              ||||||   |||      ::::::   |||:::                  :::|||
Db         16 ProLeuAlaLeu***LeuIleLeuMetAlaAlaSerGlyAlaAlaCysGluValLysAsp 35

Qy         82 GCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCA 141
                       |||   :::          :::||||||   :::|||   |||
Db         36 ValCysValGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuPro 53

Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
              |||:::|||||||||||||||   ||||||:::   |||   ||||||
Db         54 GlyArgAspGlyArgAspGly***LysGlyAspProGlyProProGlyPro--------- 70

Qy        202 GGCTTACAGGGCCCCCCTGGAAAGTTGGGG---CCTCCAGGAAAT---------CCAGGG 249
                     ||||||||||||:::         ||||||||||||         ||||||
Db         71 ------MetGlyProProGlyGluThrProCysProProGlyAsnAsnGlyLeuProGly 88

Qy        250 CCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGAT 306
                 |||   ||||||   :::|||:::||||||:::   |||   :::   |||
Db         89 AlaProGlyValProGlyGluArgGlyGluLysGlyGluAlaGlyGluArgGlyProPro 108

Qy        307 GGT---------------GATAGTAGCCTGGCTGCCTCA-----------GAAAGAAAA 339
              |||               |||      |||   |||:::                 :::
Db        109 GlyLeuProAlaHisLeuAspGluGluLeuGlnAlaThrLeuHisAspPheArgHisGln 128

Qy        340 GCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAA 399
```

```
Db        129 IleLeuGlnThrArgGlyAla--------------------------LeuSerLeuGln 139

Qy        400 --------------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                            |||||| ||| ||| ::: |||||| ::: ::: ||||||
Db        140 GlySerIleMetThrValGlyGluLysValPheSerSerAsnGlyGlnSerIleThrPhe 159

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
                  ::: :::::: ||| ::: :::||| |||||||||
Db        160 AspAlaIleGlnGluAlaCysAlaArgAlaGlyGlyArgIleAlaValProArgAsnPro 179

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
                |||||| |||||| ::: ::: ||| ||| ::: |||
Db        180 GluGluAsnGluAlaIleAlaSerPheValLysLysTyrAsnThrTyrAlaTyrValGly 199

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
                :::||::: ::: ||| ||| ||| ::: |||
Db        200 LeuThrGluGlyProSerProGlyAspPheArgTyrSerAspGlyThrProValAsnTyr 219

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
                ||||||||| ||||||||| ||| ||| |||||| :::
Db        220 ThrAsnTrpTyrArgGlyGluProAlaGlyArgGlyLys---Glu***CysValGluMet 238

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                  :::|||||||||||||||| ||| ||| ||| :::||||||
Db        239 TyrThrAspGlyGlnTrpAsnAspArgAsnCysLeuTyrSerArgLeuThrIleCysGlu 258

Qy        736 TTC 738
                |||
Db        259 Phe 259
```

RESULT 13
US-09-925-302-473
; Sequence 473, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 473
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (61)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (234)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-473

Alignment Scores:

Pred. No.:	1.77e-26	Length:	259
Score:	389.50	Matches:	95
Percent Similarity:	50.19%	Conservative:	36
Best Local Similarity:	36.40%	Mismatches:	91
Query Match:	28.53%	Indels:	39
DB:	12	Gaps:	11

US-10-054-536-2 (1-747) x US-09-925-302-473 (1-259)

```
Qy         22 CCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGAT 81
                |||||| ||| ::::::  |||::: :::|||
Db         16 ProLeuAlaLeu***LeuIleLeuMetAlaAlaSerGlyAlaAlaCysGluValLysAsp 35

Qy         82 GCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCA 141
                           |||  :::  :::|||||| :::|||  |||
Db         36 ValCysValGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuPro 53

Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
                |||:::|||||||||||||||  ||||||:::  |||  ||||||
Db         54 GlyArgAspGlyArgAspGly***LysGlyAspProGlyProProGlyPro--------- 70

Qy        202 GGCTTACAGGGCCCCCCTGGAAAGTTGGGG---CCTCCAGGAAAT---------CCAGGG 249
                      ||||||||||||:::  ||||||||||||  ||||||
Db         71 ------MetGlyProProGlyGluThrProCysProProGlyAsnAsnGlyLeuProGly 88

Qy        250 CCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGAT 306
                  |||  ||||||  :::|||:::||||||:::  |||  :::  |||
Db         89 AlaProGlyValProGlyGluArgGlyGluLysGlyGluAlaGlyGluArgGlyProPro 108

Qy        307 GGT--------------GATAGTAGCCTGGCTGCCTCA-----------GAAAGAAAA 339
                |||  |||  |||  |||:::  :::
Db        109 GlyLeuProAlaHisLeuAspGluGluLeuGlnAlaThrLeuHisAspPheArgHisGln 128

Qy        340 GCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAA 399
                ||||||||  |||  |||  |||
Db        129 IleLeuGlnThrArgGlyAla--------------------------LeuSerLeuGln 139

Qy        400 --------------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                            |||||| ||| ||| ::: |||||| ::: ::: ||||||
Db        140 GlySerIleMetThrValGlyGluLysValPheSerSerAsnGlyGlnSerIleThrPhe 159

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
                  ::: :::::: ||| ::: :::||| |||||||||
Db        160 AspAlaIleGlnGluAlaCysAlaArgAlaGlyGlyArgIleAlaValProArgAsnPro 179

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
                |||||| |||||| ::: ::: ||| ||| ::: |||
Db        180 GluGluAsnGluAlaIleAlaSerPheValLysLysTyrAsnThrTyrAlaTyrValGly 199

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
                :::||::: ::: ||| ||| ||| ::: |||
Db        200 LeuThrGluGlyProSerProGlyAspPheArgTyrSerAspGlyThrProValAsnTyr 219

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
                ||||||||| ||||||||| ||| ||| |||||| :::
Db        220 ThrAsnTrpTyrArgGlyGluProAlaGlyArgGlyLys---Glu***CysValGluMet 238

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                  :::|||||||||||||||| ||| ||| ||| :::||||||
Db        239 TyrThrAspGlyGlnTrpAsnAspArgAsnCysLeuTyrSerArgLeuThrIleCysGlu 258

Qy        736 TTC 738
                |||
Db        259 Phe 259
```

RESULT 14
US-10-312-829-10
; Sequence 10, Application US/10312829
; Publication No. US20040037781A1
; GENERAL INFORMATION:
; APPLICANT: McCormack, Francis X
; TITLE OF INVENTION: Peptides with Antioxidant and Antimicrobial Properties
; FILE REFERENCE: 10738-31
; CURRENT APPLICATION NUMBER: US/10/312,829
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: PCT/US01/21226
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,313
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10

; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-312-829-10

Alignment Scores:

Pred. No.:	9.48e-20	Length:	153
Score:	314.50	Matches:	62
Percent Similarity:	59.26%	Conservative:	18
Best Local Similarity:	45.93%	Mismatches:	52
Query Match:	23.04%	Indels:	3
DB:	12	Gaps:	1

US-10-054-536-2 (1-747) x US-10-312-829-10 (1-153)

```
Qy        343 CTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTT 402
              ||||||          ::::::   ||||||          |||      |||:::   |||
Db         19 LeuGlnAlaAlaPheSerGlnTyrLysLysValGluLeuPheProAsnGlyGlnSerVal 38

Qy        403 GGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTG 462
              |||   |||   |||   |||   |||   :::      |||   :::   :::   |||
Db         39 GlyGluLysIlePheLysThrAlaGlyPheValLysProPheThrGluAlaGlnLeuLeu 58

Qy        463 TGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATT 522
              |||   :::            :::|||:::||||||:::||||||||||||   |||:::
Db         59 CysThrGlnAlaGlyGlyGlnLeuAlaSerProArgSerAlaAlaGluAsnAlaAlaLeu 78

Qy        523 CAGAATCTCATC---------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACA 573
              |||   |||:::          |||   |||||||||   :::||||||   ||||||
Db         79 GlnGlnLeuValValAlaLysAsnGluAlaAlaPheLeuSerMetThrAspSerLysThr 98

Qy        574 GAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGT 633
              ||||||:::|||          ||||||      |||   |||:::||||||      |||
Db         99 GluGlyLysPheThrTyrProThrGlyGluSerLeuValTyrSerAsnTrpAlaProGly 118

Qy        634 GAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGG 693
              |||||||||:::   |||      ||||||||||||   :::      ||||||:::|||
Db        119 GluProAsnAspAspGlyGlySerGluAspCysValGluIlePheThrAsnGlyLysTrp 138

Qy        694 AATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
              ||||||      |||          |||   ||||||||||||
Db        139 AsnAspArgAlaCysGlyGluLysArgLeuValValCysGluPhe 153
```

RESULT 15
US-09-989-722-357
; Sequence 357, Application US/09989722
; Patent No. US20020072067A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi,Avi J.
; APPLICANT: Baker,Kevin P.
; APPLICANT: Botstein,David
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Fong,Sherman
; APPLICANT: Gerber,Hanspeter
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,J.Christopher
; APPLICANT: Gurney,Austin L.
; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C63
; CURRENT APPLICATION NUMBER: US/09/989,722
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861

; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Alignment Scores:

Pred. No.:	8.67e-19	Length:	271
Score:	304.50	Matches:	84
Percent Similarity:	44.73%	Conservative:	39
Best Local Similarity:	30.55%	Mismatches:	96
Query Match:	22.31%	Indels:	56
DB:	9	Gaps:	9

US-10-054-536-2 (1-747) x US-09-989-722-357 (1-271)

```
Qy          9 GTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGT 68
              :::|||:::   ||||||      |||:::      |||
Db         17 LeuSerLeuLeuProSerGlyHisProGlnProAlaGly---------------------- 29

Qy         69 GACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCAT 128
                                |||      ||||||  :::             ||||||::
Db         30 ----------------------AspAspAlaCysSer-ValGlnIleLeuValProGlyLe 42

Qy        129 C--------------------------AACGGCTTCCCAGGCAAAGATGGGCGTGATGG 161
              :                           |||   ||||||:::   ||||||   ||
Db         42 uLysGlyAspAlaGlyGluLysGlyAspLysGlyAlaProGlyArgProGlyArgValGl 62

Qy        162 CACCAAGGGAGAAAAGGGGGAACCAGGC---CAAGGGCTCAGAGGCTTACAGGGCCCCCC 218
              |     ||||||||||||:::   |||   :::|||   :::|||      |||
Db         62 yProThrGlyGluLysGlyAspMetGlyAspLysGlyGlnLysGlySerValGlyArgHi 82

Qy        219 TGGAAAGTTGGGGCCT---------CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACC 269
              ||||||:::||||||           |||      |||   ||||||      |||||
Db         82 sGlyLysIleGlyProIleGlySerLysGlyGluLysGlyAspSerGlyAspIleGlyPr 102

Qy        270 AAAGGGCCAAAAAGGAGACCCTGGA---------------------------------- 294
              |   |||      |||:::||||||
Db        102 oProGlyProAsnGlyGluProGlyLeuProCysGluCysSerGlnLeuArgLysAlaIl 122

Qy        295 -AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGA 353
                  |||         |||         ||||||   |||   ::::::
Db        122 eGlyGluMetAspAsnGlnValSerGlnLeuThrSerGluLeuLysPheIleLysAsnAl 142

Qy        354 AATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTT 413
              :::|||   :::                                ::::::      :::|||
Db        142 aValAlaGlyVal--------------------------ArgGluThrGluSerLysIl 153

Qy        414 CTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTT 473
              :::|||          |||         :::         :::      |||
Db        153 eTyrLeuLeuValLysGluGluLysArgTyrAlaAspAlaGlnLeuSerCysGlnGlyAr 173
```

```
Qy        474 CCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCAT 533
              :::::::::   |||::::::   |||   ||||||   :::         ::
Db        173 gGlyGlyThrLeuSerMetProLysAspGluAlaAlaAsnGlyLeuMetAlaAlaTyrLe 193

Qy        534 CAAGGAG-----------GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCA 581
              :   :::                |||:::||||||   |||   :::   ||||||
Db        193 uAlaGlnAlaGlyLeuAlaArgValPheIleGlyIleAsnAspLeuGluLysGluGlyAl 213

Qy        582 GTTTGTG------GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGA 635
               ||||||      |||   :::      |||   |||:::      |||      |||||
Db        213 aPheValTyrSerAspHisSerProMetArg---ThrPheAsnLysTrpArgSerGlyGl 232

Qy        636 ACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAA 695
              ||||||||||||||      :::||||||||||||   ::::::   :::|||   |||||
Db        232 uProAsnAsnAlaTyrAspGluGluAspCysValGluMetValAlaSerGlyGlyTrpAs 252

Qy        696 TGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
              |||||||   |||   |||:::         :::||||||||||
Db        252 nAspValAlaCysHisThrThrMetTyrPheMetCysGluPhe 266
```

Search completed: June 18, 2004, 21:13:33
Job time : 54 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 18, 2004, 20:57:11 ; Search time 21 Seconds
(without alignments)
6843.340 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 1365
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10054536/runat_18062004_173456_13600/app_query.fasta_1.903
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054536_@CGN_1_1_38_@runat_18062004_173456_13600 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1321	96.8	248	1	LNHUMC	mannose-binding le
2	790	57.9	244	1	LNMSMC	mannose-binding le
3	785	57.5	244	1	LNRTMC	mannose-binding le
4	711.5	52.1	239	1	LNMSMA	mannose-binding le
5	705.5	51.7	238	1	LNRTMA	mannose-binding le
6	454	33.3	375	1	A45225	pulmonary surfacta
7	441.5	32.3	371	1	JN0450	conglutinin precur
8	441	32.3	374	1	A42046	surfactant protein
9	440.5	32.3	371	2	I45878	conglutinin - bovi
10	418	30.6	301	2	A53570	collectin-43 - bov
11	417.5	30.6	247	1	LNRBPS	pulmonary surfacta
12	416.5	30.5	369	2	S33603	surfactant protein
13	404	29.6	248	1	LNRTPS	pulmonary surfacta
14	397.5	29.1	248	1	LNHUP1	pulmonary surfacta
15	393.5	28.8	248	1	LNHUP6	pulmonary surfacta
16	393	28.8	248	2	A48853	pulmonary surfacta
17	385.5	28.2	248	1	LNHUPS	pulmonary surfacta
18	385.5	28.2	248	2	I51921	pulmonary surfacta
19	368.5	27.0	248	1	LNDGPS	pulmonary surfacta
20	278.5	20.4	742	2	JC7595	scavenger receptor
21	272.5	20.0	277	2	JC7903	collectin liver 1
22	201.5	14.8	331	1	LNMSER	IgE Fc receptor, 1
23	199.5	14.6	207	1	LNCHL	hepatic lectin - c
24	196.5	14.4	309	1	S34198	IgE Fc receptor II
25	194.5	14.2	253	2	I49560	complement C1q B c
26	192.5	14.1	253	2	S49158	complement protein
27	192	14.1	244	2	JC4708	gelatin-binding 28
28	189	13.8	245	2	S19018	complement subcomp
29	188	13.8	1670	1	CGHU3B	collagen alpha 3(I
30	187	13.7	334	2	JC5980	ficolin-A precurso
31	187	13.7	1747	2	A45974	collagen alpha 1(X
32	187	13.7	1763	2	S16366	collagen alpha 2(I
33	187	13.7	1857	2	S31212	collagen alpha 1(X
34	187	13.7	1888	2	S78476	collagen alpha 1(X
35	186.5	13.7	245	1	C1HUQC	complement subcomp
36	186	13.6	1690	1	CGHU1B	collagen alpha 4(I
37	185.5	13.6	202	1	TTHUN	tetranectin precur
38	185.5	13.6	330	2	S46657	collagen alpha 1(X
39	185	13.6	912	2	A54423	brevican precursor
40	184	13.5	253	1	C1HUQB	complement subcomp
41	183.5	13.4	245	1	C1HUQA	complement subcomp
42	183	13.4	675	2	S20819	collagen alpha 3(I
43	183	13.4	688	2	A53330	collagen alpha 2(I
44	182.5	13.4	246	2	S29328	complement subcomp
45	182.5	13.4	673	1	CGBO6C	collagen alpha 1(I

ALIGNMENTS

RESULT 1
LNHUMC
mannose-binding lectin precursor [validated] - human
N;Alternate names: mannan-binding protein
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1989 #sequence_revision 30-Sep-1991 #text_change 08-Dec-2000
C;Accession: JL0115; S05641; A34978; JL0027; JX0319; PC2188; A32266
R;Sastry, K.; Herman, G.A.; Day, L.; Deignan, E.; Bruns, G.; Morton, C.C.; Ezekowitz, R.
J. Exp. Med. 170, 1175-1189, 1989
A;Title: The human mannose-binding protein gene. Exon structure reveals its evolutionary
A;Reference number: JL0115; MUID:90010778; PMID:2477486
A;Accession: JL0115
A;Molecule type: DNA
A;Residues: 1-248 <SAS>
A;Cross-references: EMBL:X15422; NID:g34486; PIDN:CAA33462.1; PID:g34487
R;Taylor, M.E.; Brickell, P.M.; Craig, R.K.; Summerfield, J.A.
Biochem. J. 262, 763-771, 1989
A;Title: Structure and evolutionary origin of the gene encoding a human serum mannose-b
A;Reference number: S05641; MUID:90073571; PMID:2590164
A;Accession: S05641
A;Molecule type: DNA
A;Residues: 1-248 <TAY>
A;Cross-references: EMBL:X15954; NID:g34480; PIDN:CAA34079.1; PID:g1212951
A;Accession: A34978
A;Molecule type: protein
A;Residues: 'X',22-24,'X',26,'X',28-31,'X',33-34,'X',36,'XXXX',41-50 <TAY2>
R;Ezekowitz, R.A.B.; Day, L.E.; Herman, G.A.
J. Exp. Med. 167, 1034-1046, 1988
A;Title: A human mannose-binding protein is an acute-phase reactant that shares sequence
A;Reference number: JL0027; MUID:88171281; PMID:2450948
A;Accession: JL0027
A;Molecule type: mRNA
A;Residues: 1-2,'C',4,'IT',8,'S',10-57,'R',59-60,'GT',63-106,'PGCLRK',113,'SSANRNGTYQ',
R;Kurata, H.; Sannoh, T.; Kozutsumi, Y.; Yokota, Y.; Kawasaki, T.
J. Biochem. 115, 1148-1154, 1994
A;Title: Structure and function of mannan-binding proteins isolated from human liver an
A;Reference number: JX0319; MUID:95073978; PMID:7982896
A;Accession: JX0319

A;Molecule type: protein
A;Residues: 1-248 <KUR1>
A;Accession: PC2188
A;Molecule type: mRNA
A;Residues: 1-20 <KUR2>
A;Experimental source: liver and serum
C;Comment: Mannose-binding lectins are opsonins that are important in host defense again
C;Comment: This protein is a Ca2+-requring animal lectin specific for mannose and N-acet
C;Genetics:
A;Gene: GDB:MBL
A;Cross-references: GDB:120167; OMIM:154545
A;Map position: 10q11.2-10q11.2
A;Introns: 63/1; 102/1; 125/1
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: acute phase; calcium binding; endoplasmic reticulum; Golgi apparatus; hydrox
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-248/Product: mannose-binding lectin #status experimental <MAT>
F;42-99/Region: collagen-like
F;128-244/Domain: C-type lectin homology <LCH>
F;47,73,79,82,88/Modified site: 4-hydroxyproline (Pro) (partial) #status experimental

Alignment Scores:

Pred. No.:	6.1e-110	Length:	248
Score:	1321.00	Matches:	248
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.78%	Indels:	0
DB:	1	Gaps:	0

US-10-054-536-2 (1-747) x LNHUMC (1-248)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248
```

RESULT 2
LNMSMC
mannose-binding lectin C precursor - mouse
N;Alternate names: Ra-reactive factor P28a
C;Species: Mus musculus (house mouse)
C;Date: 18-Jun-1993 #sequence_revision 20-Feb-1998 #text_change 16-Jun-2000
C;Accession: I48651; B46466; A42574; C42574
R;Sastry, R.; Wang, J.S.; Brown, D.C.; Ezekowitz, R.A.; Tauber, A.I.; Sastry, K.N.
Mamm. Genome 6, 103-110, 1995
A;Title: Characterization of murine mannose-binding protein genes Mbl1 and Mbl2 reveals
A;Reference number: I48650; MUID:95284466; PMID:7766991
A;Accession: I48651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-244 <SAS1>
A;Cross-references: EMBL:U09016; NID:g773286; PIDN:AAA82010.1; PID:g773288
R;Sastry, K.; Zahedi, K.; Lelias, J.M.; Whitehead, A.S.; Ezekowitz, R.A.
J. Immunol. 147, 692-697, 1991
A;Title: Molecular characterization of the mouse mannose-binding proteins. The mannose-
A;Reference number: A46466; MUID:91302823; PMID:1712818
A;Accession: B46466
A;Molecule type: mRNA
A;Residues: 1-2,'L',4-14,'A',16-244 <SAS2>
A;Cross-references: GB:S42294; NID:g233017; PIDN:AAB19343.1; PID:g233018
A;Experimental source: inbred CBA/J, acute phase liver library, pTZ 19 vector
A;Note: sequence extracted from NCBI backbone (NCBIN:42294, NCBIP:42295)
R;Kuge, S.; Ihara, S.; Watanabe, E.; Watanabe, M.; Takishima, K.; Suga, T.; Mamiya, G.;
Biochemistry 31, 6943-6950, 1992
A;Title: cDNAs and deduced amino acid sequences of subunits in the binding component of
A;Reference number: A42574; MUID:92345256; PMID:1637828
A;Accession: A42574
A;Molecule type: mRNA; protein
A;Residues: 1-244 <KUG>
A;Cross-references: GB:D11440; NID:g220585; PIDN:BAA02005.1; PID:g220586
A;Experimental source: BALB/c, liver
A;Note: sequence extracted from NCBI backbone (NCBIP:110137)
A;Note: parts of the sequence, including the amino end of the mature protein, were conf
A;Accession: C42574
A;Molecule type: protein
A;Residues: 19,'X',21-28,'X',30-32;72-77,'H',79-80;'G',177-185;187-189,'H',191-198 <KU2
A;Note: source is serum of ICR mice; differences may be allotypic
C;Genetics:
A;Gene: Mbl2
A;Introns: 59/1; 98/1; 121/1
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: calcium; endoplasmic reticulum; Golgi apparatus; homohexamer; hydroxyprolin
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-244/Product: mannose-binding lectin C #status experimental <MAT>
F;38-94/Region: collagen-like
F;124-240/Domain: C-type lectin homology <LCH>
F;29,34/Disulfide bonds: interchain #status predicted
F;69/Modified site: 4-hydroxyproline (Pro) #status experimental

Alignment Scores:

Pred. No.:	1.81e-62	Length:	244
Score:	790.00	Matches:	150
Percent Similarity:	75.90%	Conservative:	39
Best Local Similarity:	60.24%	Mismatches:	50
Query Match:	57.88%	Indels:	10
DB:	1	Gaps:	5

US-10-054-536-2 (1-747) x LNMSMC (1-244)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
Db          1 MetSerIlePheThrSer------PheLeuLeuLeuCysValValThrValValTyrAla 18

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
Db         19 GluThrLeuThr---GluGlyValGlnAsnSerCysPro---ValValThrCysSerSer 36

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
Db         37 ProGlyLeuAsnGlyPheProGlyLysAspGlyArgAspGlyAlaLysGlyGluLysGly 56

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTG---------GGG 231
Db         57 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysValGlyProThrGly 76

Qy        232 CCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCT 291
Db         77 ProProGlyAsnProGlyLeuLysGlyAlaValGlyProLysGlyAspArgGlyAsp--- 95

Qy        292 GGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACA 351
Db         96 ------ArgAlaGluPheAspThrSerGluIleAspSerGluIleAlaAlaLeuArgSer 113

Qy        352 GAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAG 411
Db        114 GluLeuArgAlaLeuArgAsnTrpValLeuPheSerLeuSerGluLysValGlyLysLys 133

Qy        412 TTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAG 471
Db        134 TyrPheValSerSerValLysLysMetSerLeuAspArgValLysAlaLeuCysSerGlu 153

Qy        472 TTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTC 531
Db        154 PheGlnGlySerValAlaThrProArgAsnAlaGluGluAsnSerAlaIleGlnLysVal 173

Qy        532 ATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGAT 591
Db        174 AlaLysAspIleAlaTyrLeuGlyIleThrAspValArgValGluGlySerPheGluAsp 193

Qy        592 CTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGT 651
Db        194 LeuThrGlyAsnArgValArgTyrThrAsnTrpAsnAspGlyGluProAsnAsnThrGly 213

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
Db        214 AspGlyGluAspCysValValIleLeuGlyAsnGlyLysTrpAsnAspValProCysSer 233

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
Db        234 AspSerPheLeuAlaIleCysGluPhe 242
```

RESULT 3
LNRTMC
mannose-binding lectin C precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 22-Jun-1999
C;Accession: A24791; A38322; JX0201; A26798
R;Drickamer, K.; Dordal, M.S.; Reynolds, L.
J. Biol. Chem. 261, 6878-6887, 1986
A;Title: Mannose-binding proteins isolated from rat liver contain carbohydrate-recogniti
rotein.
A;Reference number: A24791; MUID:86196130; PMID:3009480
A;Accession: A24791
A;Molecule type: mRNA
A;Residues: 1-244 <DRI>
A;Cross-references: GB:M14103
A;Note: part of the sequence, including the amino end of the mature protein, was confirm
R;Childs, R.A.; Feizi, T.; Yuen, C.T.; Drickamer, K.; Quesenberry, M.S.
J. Biol. Chem. 265, 20770-20777, 1990
A;Title: Differential recognition of core and terminal portions of oligosaccharide ligan
A;Reference number: A38322; MUID:91065871; PMID:2249985
A;Accession: A38322
A;Molecule type: mRNA
A;Residues: 86,'EL',89-97 <CHI>
R;Wada, M.; Itoh, N.; Ohta, M.; Kawasaki, T.
J. Biochem. 111, 66-73, 1992
A;Title: Characterization of rat liver mannan-binding protein gene.
A;Reference number: JX0201; MUID:92299655; PMID:1607365
A;Accession: JX0201
A;Molecule type: DNA
A;Residues: 1-244 <WAD>
A;Experimental source: liver
A;Note: the authors translated the codon CCA for residue 43 as Phe
R;Oka, S.; Itoh, N.; Kawasaki, T.; Yamashina, I.
J. Biochem. 101, 135-144, 1987
A;Title: Primary structure of rat liver mannan-binding protein deduced from its cDNA seq
A;Reference number: A26798; MUID:87194686; PMID:3032924
A;Accession: A26798
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-244 <OKA>
A;Cross-references: GB:X05023; NID:g56634; PIDN:CAA28687.1; PID:g56635
C;Comment: The molecule contains six identical chains, occurring as disulfide-bonded dim
C;Comment: Mannan-binding proteins are calcium ion-dependent and are specific for mannos
C;Genetics:
A;Introns: 59/1; 98/1; 121/1
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: calcium; endoplasmic reticulum; Golgi apparatus; homohexamer; hydroxyproline
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-244/Product: mannose-binding lectin C #status experimental <MAT>
F;38-94/Region: collagen-like
F;124-240/Domain: C-type lectin homology <LCH>
F;29,34/Disulfide bonds: interchain #status predicted
F;69/Modified site: 4-hydroxyproline (Pro) #status experimental

Alignment Scores:

Pred. No.:	5.06e-62	Length:	244
Score:	785.00	Matches:	151
Percent Similarity:	75.20%	Conservative:	34
Best Local Similarity:	61.38%	Mismatches:	57
Query Match:	57.51%	Indels:	4
DB:	1	Gaps:	3

US-10-054-536-2 (1-747) x LNRTMC (1-244)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
Db          1 MetSerLeuPheThrSer------PheLeuLeuLeuCysValLeuThrAlaValTyrAla 18

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
Db         19 GluThrLeuThr---GluGlyAlaGlnSerSerCysPro---ValIleAlaCysSerSer 36

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
Db         37 ProGlyLeuAsnGlyPheProGlyLysAspGlyHisAspGlyAlaLysGlyGluLysGly 56

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
Db         57 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysValGlyProAlaGly 76

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
Db         77 ProProGlyAsnProGlySerLysGlyAlaThrGlyProLysGlyAspArgGlyGluSer 96

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
Db         97 ValGluPheAspThrThrAsnIleAspLeuGluIleAlaAlaLeuArgSerGluLeuArg 116

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
Db        117 AlaMetArgLysTrpValLeuLeuSerMetSerGluAsnValGlyLysLysTyrPheMet 136

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
```

```
Db        137 SerSerValArgArgMetProLeuAsnArgAlaLysAlaLeuCysSerGluLeuGlnGly 156

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              :::|||||||||||||||||||||||    ||||||   ||||||||||||:::   |||:::
Db        157 ThrValAlaThrProArgAsnAlaGluGluAsnArgAlaIleGlnAsnValAlaLysAsp 176

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              |||||||||||||||||||||||:::::|||||      |||   ||||||||||||
Db        177 ValAlaPheLeuGlyIleThrAspGlnArgThrGluAsnValPheGluAspLeuThrGly 196

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||:::   |||||||||||||||||||||||||||||||||   ||||||   |||
Db        197 AsnArgValArgTyrThrAsnTrpAsnGluGlyGluProAsnAsnValGlySerGlyGlu 216

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              :::||||||:::||||||   ||||||:::|||||||||||||||||||   |||
Db        217 AsnCysValValLeuLeuThrAsnGlyLysTrpAsnAspValProCysSerAspSerPhe 236

Qy        721 CTGGCCGTCTGTGAGTTC 738
              |||   ||||||||||||
Db        237 LeuValValCysGluPhe 242
```

RESULT 4
LNMSMA
mannose-binding lectin A precursor - mouse
N;Alternate names: Ra-reactive factor P28b; serum mannan-binding protein
C;Species: Mus musculus (house mouse)
C;Date: 18-Jun-1993 #sequence_revision 02-Jun-1994 #text_change 16-Jun-2000
C;Accession: A46466; B42574; I48650
R;Sastry, K.; Zahedi, K.; Lelias, J.M.; Whitehead, A.S.; Ezekowitz, R.A.
J. Immunol. 147, 692-697, 1991
A;Title: Molecular characterization of the mouse mannose-binding proteins. The mannose-b
A;Reference number: A46466; MUID:91302823; PMID:1712818
A;Accession: A46466
A;Molecule type: mRNA
A;Residues: 1-239 <SAS1>
A;Cross-references: GB:S42292; NID:g1679939; PIDN:AAB19342.1; PID:g233016
A;Experimental source: inbred CBA/J, acute phase liver library, pTZ 19 vector
A;Note: sequence extracted from NCBI backbone (NCBIN:42292, NCBIP:42293)
R;Kuge, S.; Ihara, S.; Watanabe, E.; Watanabe, M.; Takishima, K.; Suga, T.; Mamiya, G.;
Biochemistry 31, 6943-6950, 1992
A;Title: cDNAs and deduced amino acid sequences of subunits in the binding component of
A;Reference number: A42574; MUID:92345256; PMID:1637828
A;Accession: B42574
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA; protein
A;Residues: 1-239 <KUG>
A;Cross-references: GB:D11441; NID:g220587; PIDN:BAA02006.1; PID:g220588
A;Experimental source: BALB/c, liver
A;Note: sequence extracted from NCBI backbone (NCBIP:110142)
R;Sastry, R.; Wang, J.S.; Brown, D.C.; Ezekowitz, R.A.; Tauber, A.I.; Sastry, K.N.
Mamm. Genome 6, 103-110, 1995
A;Title: Characterization of murine mannose-binding protein genes Mbl1 and Mbl2 reveals
A;Reference number: I48650; MUID:95284466; PMID:7766991
A;Accession: I48650
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-239 <SAS2>
A;Cross-references: EMBL:U09010; NID:g773278; PIDN:AAA82009.1; PID:g773280
C;Comment: Mannose-binding lectins are opsonins that are important in host defense again
C;Genetics:
A;Gene: Mbl1
A;Introns: 52/1; 91/1; 116/1
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: acute phase; calcium; hydroxyproline; lectin; liver; plasma
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-238/Product: mannose-binding lectin A #status predicted <MAT>
F;36-88/Region: collagen-like
F;119-235/Domain: C-type lectin homology <LCH>

Alignment Scores:
Pred. No.: 1.88e-55 Length: 239
Score: 711.50 Matches: 144
Percent Similarity: 69.48% Conservative: 29
Best Local Similarity: 57.83% Mismatches: 63
Query Match: 52.12% Indels: 13
DB: 1 Gaps: 4

US-10-054-536-2 (1-747) x LNMSMA (1-239)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||   |||   |||   ||||||   :::||||||   :::|||:::   |||   |||
Db          1 MetLeuLeuLeuProLeuLeuPro---ValLeuLeuCysValValSerValSerSerSer 19

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
                 :::   ||||||||||||      |||||||||   :::||||||||||||
Db         20 GlySerGlnThrCysGluAspThrLeuLysThrCys---SerValIleAlaCys------ 36

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
                                   |||:::||||||||||||||   ||||||||||||||
Db         37 --------------------GlyArgAspGlyArgAspGlyProLysGlyGluLysGly 49

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 69

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              :::   |||      |||||||||||||||||||||||||||||||   |||   :::
Db         70 SerValGlySerProGlySerProGlyProLysGlyGlnLysGlyAspHisGlyAspAsn 89

Qy        301 CCGGATGGTGATAGTAGCCTGGCT------GCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
                    :::     ||||||        :::|||   :::   |||:::::::::
Db         90 ArgAlaIleGluGluLysLeuAlaAsnMetGluAlaGluIleArgIleLeuLysSerLys 109

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
              :::          |||        ||||||:::|||||||:::   |||   |||
Db        110 LeuGlnLeuThrAsnLysLeuHisAlaPheSerMetGlyLysLysSerGlyLysLysLeu 129

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              |||:::||||||   |||   |||   |||   |||||||||:::||||||   :::
Db        130 PheValThrAsnHisGluLysMetProPheSerLysValLysSerLeuCysThrGluLeu 149

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
              |||   :::||||||   ||||||||||||   ||||||   |||||||||   :::
Db        150 GlnGlyThrValAlaIleProArgAsnAlaGluGluAsnLysAlaIleGlnGluValAla 169

Qy        535 AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTG 594
                      ||||||||||||||||||||||||   ||||||||||||||:::   :::
Db        170 ThrGlyIleAlaPheLeuGlyIleThrAspGluAlaThrGluGlyGlnPheMetTyrVal 189

Qy        595 ACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCT 654
              ||||||   ||||||||||||:::||||||   :::   ||||||||||||   ||||||
Db        190 ThrGlyGlyArgLeuThrTyrSerAsnTrpLysLysAspGluProAsnAsnHisGlySer 209

Qy        655 GATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACC 714
                 ||||||||||||::::::|||   ||||||   |||||||||:::   |||
Db        210 GlyGluAspCysValIleIleLeuAspAsnGlyLeuTrpAsnAspIleSerCysGlnAla 229

Qy        715 TCCCATCTGGCCGTCTGTGAGTTCCCT 741
              |||      ||||||||||||||||||
Db        230 SerPheLysAlaValCysGluPhePro 238
```

RESULT 5
LNRTMA
mannose-binding lectin A precursor - rat
N;Alternate names: serum mannan-binding protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 22-Jun-1999
C;Accession: B24791; A29556; A27799
R;Drickamer, K.; Dordal, M.S.; Reynolds, L.
J. Biol. Chem. 261, 6878-6887, 1986
A;Title: Mannose-binding proteins isolated from rat liver contain carbohydrate-recognit
rotein.
A;Reference number: A24791; MUID:86196130; PMID:3009480
A;Accession: B24791
A;Molecule type: mRNA

```
A;Residues: 1-238 <DR1>
R;Drickamer, K.; McCreary, V.
J. Biol. Chem. 262, 2582-2589, 1987
A;Title: Exon structure of a mannose-binding protein gene reflects its evolutionary rela
A;Reference number: A29556; MUID:87137502; PMID:3029088
A;Accession: A29556
A;Molecule type: DNA
A;Residues: 1-155,'K',157-238 <DR2>
A;Cross-references: GB:M14104; GB:M14105; NID:g205259; PIDN:AAA98781.1; PID:g205261
A;Note: the codon AAG for residue 156 is inconsistent with the authors' statement that t
R;Ikeda, K.; Sannoh, T.; Kawasaki, N.; Kawasaki, T.; Yamashina, I.
J. Biol. Chem. 262, 7451-7454, 1987
A;Title: Serum lectin with known structure activates complement through the classical pa
A;Reference number: A27799; MUID:87222358; PMID:3584121
A;Accession: A27799
A;Molecule type: protein
A;Residues: 18-42 <IKE>
C;Comment: Mannose-binding lectins are opsonins that are important in host defense again
C;Comment: This plasma protein binds mannose and N-acetylglucosamine and can activate cc
C;Comment: The molecule consists of approximately 20 identical chains linked by disulfid
C;Genetics:
A;Introns: 51/1; 90/1; 115/1
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: acute phase; calcium; hydroxylysine; hydroxyproline; lectin; liver; plasma
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-238/Product: mannose-binding lectin A #status experimental <MAT>
F;36-88/Region: collagen-like
F;85-87/Region: cell attachment (R-G-D) motif
F;118-234/Domain: C-type lectin homology <LCH>
F;61,67,73/Modified site: 4-hydroxyproline (Pro) #status experimental
F;79,82/Modified site: lysine derivative (Lys) (probably 5-hydroxylysine) #status experi

Alignment Scores:
Pred. No.:                6.47e-55      Length:          238
Score:                    705.50        Matches:         143
Percent Similarity:       70.08%        Conservative:    28
Best Local Similarity:    58.61%        Mismatches:      60
Query Match:              51.68%        Indels:          13
DB:                       1             Gaps:            4

US-10-054-536-2 (1-747) x LNRTMA (1-238)

Qy          19 CTCCCTCTCCTT---CTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGT 75
               ||||||||||||   ||||||   :::|||:::   |||   |||   :::   ||||||
Db           4 LeuProLeuLeuValLeuLeuCysValValSerValSerSerSerGlySerGlnThrCys 23

Qy          76 GAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGC 135
               |||:::      |||||||||   :::||||||||||||
Db          24 GluGluThrLeuLysThrCys---SerValIleAlaCys---------------------- 35

Qy         136 TTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGG 195
                  |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||
Db          36 ------GlyArgAspGlyArgAspGlyProLysGlyGluLysGlyGluProGlyGlnGly 53

Qy         196 CTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCT 255
               |||||||||||||||||||||||||||||||||||||||||||:::   |||
Db          54 LeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGlySerValGlyAlaPro 73

Qy         256 GGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGT 315
               ||||||   ||||||||||||||||||||||   |||   |||          :::
Db          74 GlySerGlnGlyProLysGlyGlnLysGlyAspArgGlyAspSerArgAlaIleGluVal 93

Qy         316 AGCCTGGCT------GCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAA 369
                  ||||||         :::|||         |||:::::::::::::
Db          94 LysLeuAlaAsnMetGluAlaGluIleAsnThrLeuLysSerLysLeuGluLeuThrAsn 113

Qy         370 AAGTGGCTGACCTTCTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGT 429
               |||          ||||||:::||||||:::   |||   |||||||||:::||||||
Db         114 LysLeuHisAlaPheSerMetGlyLysLysSerGlyLysLysPhePheValThrAsnHis 133

Qy         430 GAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCC 489
               |||   |||   |||   ||||||||||||||||   :::   :::   :::||||||
Db         134 GluArgMetProPheSerLysValLysAlaLeuCysSerGluLeuArgGlyThrValAla 153

Qy         490 ACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTC 549
               ||||||||||||   ||||||   ||||||||   :::   |||      ||||||
Db         154 IleProArgAsnAlaGluGluAsnLysAlaIleGlnGluValAlaLysThrSerAlaPhe 173

Qy         550 CTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTG 609
               ||||||||||||||||||   ||||||||||||||:::   :::||||||   ||||||
Db         174 LeuGlyIleThrAspGluValThrGluGlyGlnPheMetTyrValThrGlyGlyArgLeu 193

Qy         610 ACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTA 669
               ||||||:::||||||   :::   |||||||||:::   ||||||   ||||||||||||
Db         194 ThrTyrSerAsnTrpLysLysAspGluProAsnAspHisGlySerGlyGluAspCysVal 213

Qy         670 TTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTC 729
                  ::::::   ||||||   |||||||||:::   |||      ||||||   ||||||
Db         214 ThrIleValAspAsnGlyLeuTrpAsnAspIleSerCysGlnAlaSerHisThrAlaVal 233

Qy         730 TGTGAGTTCCCT 741
               ||||||||||||
Db         234 CysGluPhePro 237

RESULT 6
A45225
pulmonary surfactant protein D precursor - human
N;Alternate names: SP-D
C;Species: Homo sapiens (man)
C;Date: 16-Apr-1999 #sequence_revision 16-Apr-1999 #text_change 22-Jun-1999
C;Accession: A45225; S23434; S24555; S44420; S18382; A56776
R;Crouch, E.; Rust, K.; Veile, R.; Donis-Keller, H.; Grosso, L.
J. Biol. Chem. 268, 2976-2983, 1993
A;Title: Genomic organization of human surfactant protein D (SP-D). SP-D is encoded on c
A;Reference number: A45225; MUID:93155122; PMID:8428971
A;Accession: A45225
A;Molecule type: DNA
A;Residues: 1-375 <CRO>
A;Cross-references: GB:L05483; GB:L05484; GB:L05485; NID:g292505; PIDN:AAB59450.1; PID:g
A;Experimental source: placenta
A;Note: sequence extracted from NCBI backbone (NCBIP:124316)
R;Lu, J.; Willis, A.C.; Reid, K.B.M.
Biochem. J. 284, 795-802, 1992
A;Title: Purification, characterization and cDNA cloning of human lung surfactant protei
A;Reference number: S23434; MUID:92322003; PMID:1339284
A;Accession: S23434
A;Molecule type: mRNA
A;Residues: 1-30,'T',32-121,'P',123-179,'A',181-375 <LUJ1>
A;Cross-references: EMBL:X65018; NID:g34766; PIDN:CAA46152.1; PID:g34767
A;Experimental source: lung
A;Accession: S24555
A;Molecule type: protein
A;Residues: 214-234,'X',236,'XX',239-241 <LUJ2>
R;Hoppe, H.J.; Barlow, P.N.; Reid, K.B.M.
FEBS Lett. 344, 191-195, 1994
A;Title: A parallel three stranded alpha-helical bundle at the nucleation site of collag
A;Reference number: S44420; MUID:94244769; PMID:8187882
A;Accession: S44420
A;Molecule type: mRNA
A;Residues: 202-257 <HOP>
R;Rust, K.; Grosso, L.; Zhang, V.; Chang, D.; Persson, A.; Longmore, W.; Cai, G.Z.; Crou
Arch. Biochem. Biophys. 290, 116-126, 1991
A;Title: Human surfactant protein D: SP-D contains a C-type lectin carbohydrate recognit
A;Reference number: S18382; MUID:91378578; PMID:1898081
A;Accession: S18382
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 'F',60-205,'P',207-374,'HF' <RUS>
A;Cross-references: GB:L05485; NID:g292505
A;Note: corrections to this sequence are reported in reference A56776
R;Crouch, E.; Persson, A.; Chang, D.
Am. J. Pathol. 142, 241-248, 1993
A;Title: Accumulation of surfactant protein D in human pulmonary alveolar proteinosis.
A;Reference number: A56776; MUID:93142849; PMID:8424457
A;Accession: A56776
A;Status: preliminary
```

```
A;Molecule type: protein
A;Residues: 46-58,'F',60-62,'E',64-72;223-227,'X',229-239,'P',241-245,'X',247-256,'X',25
A;Cross-references: PIDN:AAB25037.1; PID:g263973; PIDN:AAB25038.1; PID:g263974
A;Experimental source: bronchoalveolar lavage
A;Note: sequence extracted from NCBI backbone (NCBIP:123024, NCBIP:123023); sequence mod
C;Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t
C;Comment: This protein is synthesized by alveolar type II cells.
C;Genetics:
A;Gene: GDB:SFTPD; SFTP4; SP-D
A;Cross-references: GDB:132674; OMIM:178635
A;Map position: 10q22.2-10q23.1
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
C;Keywords: blocked amino end; calcium; glycoprotein; hydroxylysine; hydroxyproline; lun
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-375/Product: pulmonary surfactant protein D #status predicted <MAT>
F;21-45/Domain: non-collagenous #status predicted <NC1>
F;46-222/Domain: collagenous #status predicted <COL>
F;223-375/Domain: non-collagenous #status predicted <NC2>
F;254-373/Domain: C-type lectin homology <LCH>
F;90/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;281-373,351-365/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.:                2.09e-32      Length:          375
Score:                    454.00        Matches:         99
Percent Similarity:       55.07%        Conservative:    26
Best Local Similarity:    43.61%        Mismatches:      82
Query Match:              33.26%        Indels:          20
DB:                       1             Gaps:            5

US-10-054-536-2 (1-747) x A45225 (1-375)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||:::   |||      |||      |||      |||   |||||||||:::
Db        149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||      |||      |||      |||
Db        169 GlyValProGlyGluArgGlyValProGlyAsnThrGlyAlaAlaGlySerAlaGlyAla 188

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||:::||||||      |||   ||||||   ||||||   ||||||
Db        189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy        286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
                 ||||||   |||         |||:::|||   |||         ||||||   |||
Db        209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy        337 AAAGCT----------------------------CTGCAAACAGAAATGGCACGTATC 366
              :::                              ||||||       ::::::
Db        229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||      |||      |||:::   ||||||   |||   |||   |||
Db        249 LysLysValGluLeuPheProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::      |||   :::   :::   ||||||   :::          :::
Db        269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
              |||:::||||||:::||||||||||||   |||:::|||   |||:::
Db        289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              |||   ||||||||   :::||||||   ||||||||||||:::|||          |||
Db        309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
              |||      |||   |||:::||||||      ||||||||||||:::   |||
Db        329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySer 348

Qy        658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              |||||||||||   :::      ||||||:::||||||||      |||
Db        349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy        718 CATCTGGCCGTCTGTGAGTTC 738
                  |||   ||||||||||||
Db        369 ArgLeuValValCysGluPhe 375

RESULT 7
JN0450
conglutinin precursor - bovine
N;Alternate names: C3b-binding protein
N;Contains: conglutinin-N
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JN0450; JC2396; S33235; A23740; S36879; S35044; I46010; A29416; S34054
R;Suzuki, Y.; Yin, Y.P.; Makino, M.; Kurimura, T.; Wakamiya, N.
Biochem. Biophys. Res. Commun. 191, 335-342, 1993
A;Title: Cloning and sequencing of a cDNA coding for bovine conglutinin.
A;Reference number: JN0450; MUID:93213261; PMID:8460993
A;Accession: JN0450
A;Molecule type: mRNA
A;Residues: 1-371 <SUZ>
A;Cross-references: DDBJ:D14085; NID:g285643; PIDN:BAA03170.1; PID:g285644
A;Experimental source: liver
R;Kawasaki, N.; Itoh, N.; Kawasaki, T.
Biochem. Biophys. Res. Commun. 198, 597-604, 1994
A;Title: Gene organization and 5'-flanking region sequence of conglutinin: A C-type mam
A;Reference number: JC2396; MUID:94128104; PMID:8297370
A;Accession: JC2396
A;Molecule type: mRNA
A;Residues: 1-371 <KA2>
A;Note: The authors translated the codon GAT for residues 250 and 270 as Glu
R;Lu, J.; Laursen, S.B.; Thiel, S.; Jensenius, J.C.; Reid, K.B.M.
Biochem. J. 292, 157-162, 1993
A;Title: The cDNA cloning of conglutinin and identification of liver as a primary site
A;Reference number: S33235; MUID:93277452; PMID:7684896
A;Accession: S33235
A;Molecule type: mRNA
A;Residues: 1-172,'H',174-217,'A',219-271,'V',273-371 <LUJ>
A;Cross-references: EMBL:X71774; NID:g395267; PIDN:CAA50665.1; PID:g395268
A;Experimental source: liver
R;Lee, Y.M.; Leiby, K.R.; Allar, J.; Paris, K.; Lerch, B.; Okarma, T.B.
J. Biol. Chem. 266, 2715-2723, 1991
A;Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin
A;Reference number: A23740; MUID:91131556; PMID:1993651
A;Accession: A23740
A;Molecule type: protein
A;Residues: 21-209,'S',211-371 <LEE>
R;Kawasaki, N.; Yokota, Y.; Kawasaki, T.
Arch. Biochem. Biophys. 305, 533-540, 1993
A;Title: Differentiation of conglutination activity and sugar-binding activity of congl
A;Reference number: S36879; MUID:93384312; PMID:8373191
A;Accession: S36879
A;Molecule type: protein
A;Residues: 21-54;75-86,'X',88-89,'X',91,'X',93-94;208-209,'X',211-227 <KAW>
A;Experimental source: serum
R;Lu, J.; Wiedemann, H.; Holmskov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.
Eur. J. Biochem. 215, 793-799, 1993
A;Title: Structural similarity between lung surfactant protein D and conglutinin. Two d
A;Reference number: S35044; MUID:93358905; PMID:8354286
A;Accession: S35044
A;Molecule type: protein
A;Residues: 75-86,'X',88-89,'X',91,'I' <LUA>
A;Experimental source: lung
R;Young, N.M.; Leon, M.A.
Biochem. Biophys. Res. Commun. 143, 645-651, 1987
A;Title: The carbohydrate specificity of conglutinin and its homology to proteins in th
A;Reference number: A29416; MUID:87184551; PMID:3566740
A;Contents: annotation
R;Malhotra, R.; Laursen, S.B.; Willis, A.C.; Sim, R.B.
Biochem. J. 293, 15-19, 1993
A;Title: Research Communication. Localization of the receptor-binding site in the colle
A;Reference number: S34054; MUID:93319501; PMID:8328957
```

```
A;Contents: annotation
R;Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry
J. Immunol. 153, 173-180, 1994
A;Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship to
A;Reference number: I46010; MUID:94267222; PMID:8207234
A;Accession: I46010
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-371 <LIO>
A;Cross-references: EMBL:U06860; NID:g507183; PIDN:AAB60624.1; PID:g514256
C;Comment: This protein mediates the agglutination of erythrocytes with antibody and com
C;Comment: This protein is a Ca2+-dependent serum lectin specific for N-acetylglucosamin
C;Genetics:
A;Gene: CGN1
A;Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
C;Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-371/Product: conglutinin #status predicted <MAT>
F;46-214/Region: collagen-like
F;75-371/Product: conglutinin-N #status predicted <MA2>
F;248-369/Domain: C-type lectin homology <LCH>
F;63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #status p
F;63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxylysine (Lys) #status experime
F;78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxyproline (Pro) #status ex
F;337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Alignment Scores:
Pred. No.:                2.73e-31      Length:          371
Score:                    441.50        Matches:         91
Percent Similarity:       56.77%        Conservative:    39
Best Local Similarity:    39.74%        Mismatches:      74
Query Match:              32.34%        Indels:          25
DB:                       1             Gaps:            6

US-10-054-536-2 (1-747) x JN0450 (1-371)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::|||||||||   ||||||||||||                |||   ||||||||||||
Db        146 AlaProGlyIleGlnGlyPheProGlyPro---------SerGlyLeuLysGlyGluLys 162

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||      |||:::   ||||||   |||
Db        163 GlyAlaProGlyGluThrGlyAlaProGlyArgAlaGlyValThrGlyProSerGlyAla 182

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||      |||      |||   ||||||   ||||||   :::|||
Db        183 IleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGly 202

Qy        286 GACCCTGGAAAAAGT---CCGGATGGTGATAGTAGCCTGGCTGCC-------------- 327
              |||||||||::::::         |||:::|||   ||||||
Db        203 AspProGlyGluThrGlyAlaLysGlyGluSerGlyLeuAlaGluValAsnAlaLeuLys 222

Qy        328 --------------------TCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
                                  :::     |||      ::::::
Db        223 GlnArgValThrIleLeuAspGlyHisLeuArgArgPheGlnAsnAlaPheSerGlnTyr 242

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||   :::   |||      |||:::   ||||||   |||   |||   |||
Db        243 LysLysAlaValLeuPheProAspGlyGlnAlaValGlyGluLysIlePheLysThrAla 262

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::   ::::::         :::   ||||||   :::   :::      :::
Db        263 GlyAlaValLysSerTyrSerAspAlaGluGlnLeuCysArgGluAlaLysGlyGlnLeu 282

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG--------- 537
              |||:::||||||:::::|||||||||   |||:::      :::::::::
Db        283 AlaSerProArgSerSerAlaGluAsnGluAlaValThrGlnMetValArgAlaGlnGlu 302

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              :::   |||:::|||   :::   |||      |||||||||:::|||          |||
Db        303 LysAsnAlaTyrLeuSerMetAsnAspIleSerThrGluGlyArgPheThrTyrProThr 322

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT------GGT 651
              |||      |||   |||:::||||||   :::|||||||||||||||:::      |||
Db        323 GlyGluIleLeuValTyrSerAsnTrpAlaAspGlyGluProAsnAsnSerAspGluGly 342

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
                  |||:::||||||   :::      :::|||:::||||||||||||||||||||
Db        343 GlnProGluAsnCysValGluIlePheProAspGlyLysTrpAsnAspValProCysSer 362

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
                    |||   :::|||||||||
Db        363 LysGlnLeuLeuValIleCysGluPhe 371

RESULT 8
A42046
surfactant protein D - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A42046
R;Shimizu, H.; Fisher, J.H.; Papst, P.; Benson, B.; Lau, K.; Mason, R.J.; Voelker, D.R.
J. Biol. Chem. 267, 1853-1857, 1992
A;Title: Primary structure of rat pulmonary surfactant protein D. cDNA and deduced amino
A;Reference number: A42046; MUID:92112913; PMID:1370483
A;Accession: A42046
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-374 <SHI>
A;Cross-references: GB:M81231; NID:g207035; PIDN:AAA42170.1; PID:g207036
A;Experimental source: lung
A;Note: sequence extracted from NCBI backbone (NCBIN:76027, NCBIP:76031)
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
F;253-372/Domain: C-type lectin homology <LCH>

Alignment Scores:
Pred. No.:                3.03e-31      Length:          374
Score:                    441.00        Matches:         110
Percent Similarity:       38.81%        Conservative:    34
Best Local Similarity:    29.65%        Mismatches:      99
Query Match:              32.31%        Indels:          128
DB:                       1             Gaps:            7

US-10-054-536-2 (1-747) x A42046 (1-374)

Qy         10 TTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTG 69
              |||   |||:::   :::||||||:::      :::              :::|||
Db          4 PheLeuSerMetLeuValLeuLeuValGlnProLeuGlyAspLeuGlyAlaGluMetLys 23

Qy         70 ACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATC 129
              |||                     |||          :::          ||||||
Db         24 ThrLeuSerGlnArgSerIleThrAsnThrCysThrLeuValLeuCysSerProThrGlu 43

Qy        130 AACGGCTTCCCAGGCAAAGATGGGCGTGAT---------GGCACCAAGGGAGAAAAGGGG 180
              ||||||   ||||||:::||||||||||||         |||   :::||||||||||||
Db         44 AsnGlyLeuProGlyArgAspGlyArgAspGlyArgGluGlyProArgGlyGluLysGly 63

Qy        181 GAACCAGGC---------------------------------------------------- 189
              :::||||||
Db         64 AspProGlyLeuProGlyProMetGlyLeuSerGlyLeuProGlyProArgGlyProVal 83

Qy        190 ----------------------------------------CAAGGGCTCAGAGGCTTA 207
                                                     :::|||   |||||||||
Db         84 GlyProLysGlyGluAsnGlySerAlaGlyGluProGlyProLysGlyGluArgGlyLeu 103

Qy        208 CAGGGCCCCCCT---------------------------------------------- 219
                 |||||||||
Db        104 ValGlyProProGlySerProGlyIleSerGlyProAlaGlyLysGluGlyProSerGly 123

Qy        220 ------GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAG 273
                    |||   :::||||||   |||   |||||||||   |||      |||||||||
Db        124 LysGlnGlyAsnIleGlyProGlnGlyLysProGlyProLysGlyGluAlaGlyProLys 143

Qy        274 GGCCAAAAAGGAGACCCTGGA--------------------AAAAGTCCGGATGGT--- 309
              |||:::   |||   ||||||                        |||   |||   |||
Db        144 GlyGluValGlyAlaProGlyMetGlnGlySerAlaGlyAlaLysGlyProAlaGlyPro 163
```

```
Qy        309 ------------------------------------------------------------ 309
Db        164 LysGlyGluArgGlyAlaProGlyGluGlnGlyAlaProGlyAsnAlaGlyAlaAlaGly 183

Qy        309 ------------------------------------------------------------ 309
Db        184 ProAlaGlyProAlaGlyProGlnGlyAlaProGlySerArgGlyProProGlyLeuLys 203

Qy        310 ------------------------------------------------------GATAGT 315
                                                                      ||||||
Db        204 GlyAspArgGlyAlaProGlyAspArgGlyIleLysGlyGluSerGlyLeuProAspSer 223

Qy        316 AGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC--------- 366
              :::             :::   :::||||||      ::::::   |||:::
Db        224 AlaAlaLeuArgGlnGlnMetGluAlaLeuAsnGlyLysLeuGlnArgLeuGluAlaAla 243

Qy        367 ------------AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
                          ||||||      |||      |||:::   ||||||:::|||
Db        244 PheSerArgTyrLysLysAlaAlaLeuProAspGlyGlnSerValGlyAspLysIle 263

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              |||                         ||||||      |||   :::|||   :::
Db        264 PheArgAlaAlaAsnSerGluGluProPheGluAspAlaLysGluMetCysArgGlnAla 283

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
                  :::|||:::||||||:::|||   ||||||   |||:::|||   |||:::
Db        284 GlyGlyGlnLeuAlaSerProArgSerAlaThrGluAsnAlaAlaValGlnGlnLeuVal 303

Qy        535 ---------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTT 585
                      :::   |||||||||   :::||||||      ||||||||||:::|||
Db        304 ThrAlaHisSerLysAlaAlaPheLeuSerMetThrAspValGlyThrGluGlyLysPhe 323

Qy        586 GTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAAT 645
                    ||||||      |||   |||:::||||||      ||||||||||||||||
Db        324 ThrTyrProThrGlyGluAlaLeuValTyrSerAsnTrpAlaProGlyGluProAsnAsn 343

Qy        646 GCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCC 705
                |||      |||:::||||||   :::      ||||||||||||||||||
Db        344 AsnGlyGlyAlaGluAsnCysValGluIlePheThrAsnGlyGlnTrpAsnAspLysAla 363

Qy        706 TGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
              |||           |||   :::||||||||||
Db        364 CysGlyGluGlnArgLeuValIleCysGluPhe 374
```

RESULT 9
I45878
conglutinin - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 20-Aug-1999
C;Accession: I45878
R;Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry
Gene 141, 277-281, 1994
A;Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization c
A;Reference number: I45878; MUID:94215917; PMID:8163202
A;Accession: I45878
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-371 <LIO>
A;Cross-references: GB:L18871; NID:g495012; PIDN:AAA20126.1; PID:g495013
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
F;248-369/Domain: C-type lectin homology <LCH>

Alignment Scores:

Pred. No.:	3.35e-31	Length:	371
Score:	440.50	Matches:	91
Percent Similarity:	56.77%	Conservative:	39
Best Local Similarity:	39.74%	Mismatches:	74
Query Match:	32.27%	Indels:	25
DB:	2	Gaps:	6

US-10-054-536-2 (1-747) x I45878 (1-371)

```
Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::|||||||||   ||||||||||||                |||   ||||||||||||
Db        146 AlaProGlyIleGlnGlyPheProGlyPro---------SerGlyLeuLysGlyGluLys 162

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::             |||     |||:::   ||||||   |||
Db        163 GlyAlaProGlyGluThrGlyAlaProGlyHisAlaGlyValThrGlyProSerGlyAla 182

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||     |||     |||   ||||||   ||||||   :::|||
Db        183 IleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGly 202

Qy        286 GACCCTGGAAAAAGT---CCGGATGGTGATAGTAGCCTGGCTGCC--------------- 327
              |||||||||:::::         |||:::|||   ||||||
Db        203 AspProGlyGluThrGlyAlaLysGlyGluSerGlyLeuAlaGluValAsnAlaLeuLys 222

Qy        328 --------------------TCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
                                      :::     |||         ::::::
Db        223 GlnArgValThrIleLeuAspGlyHisLeuArgArgPheGlnAsnAlaPheSerGlnTyr 242

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||   :::   |||     |||:::   ||||||   |||   |||   |||
Db        243 LysLysAlaValLeuPheProAspGlyGlnAlaValGlyGluLysIlePheLysThrAla 262

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::   ::::::         :::   ||||||   :::   :::       :::
Db        263 GlyAlaValLysSerTyrSerAspAlaGluGlnLeuCysArgGluAlaLysGlyGlnLeu 282

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG--------- 537
              |||:::||||||:::::|||||||||   |||:::      :::::::::
Db        283 AlaSerProArgSerSerAlaGluAsnGluAlaValThrGlnMetValArgAlaGlnGlu 302

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              :::   |||:::|||   :::   |||       |||||||||:::|||         |||
Db        303 LysAsnAlaTyrLeuSerMetAsnAspIleSerThrGluGlyArgPheThrTyrProThr 322

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT------GGT 651
              |||      |||   |||:::||||||   :::|||||||||||||||:::      |||
Db        323 GlyGluIleLeuValTyrSerAsnTrpAlaAspGlyGluProAsnAsnSerAspGluGly 342

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
                    |||:::||||||   :::      :::|||:::|||||||||||||||||||||
Db        343 GlnProGluAsnCysValGluIlePheProAspGlyLysTrpAsnAspValProCysSer 362

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
                    |||   :::||||||||||
Db        363 LysGlnLeuLeuValIleCysGluPhe 371
```

RESULT 10
A53570
collectin-43 - bovine
N;Alternate names: lectin CL-43
C;Species: Bos primigenius taurus (cattle)
C;Date: 12-Apr-1995 #sequence_revision 23-Feb-1996 #text_change 17-Mar-1999
C;Accession: A53570; A46689
R;Lim, B.L.; Willis, A.C.; Reid, K.B.M.; Lu, J.; Laursen, S.B.; Jensenius, J.C.; Holmsko
J. Biol. Chem. 269, 11820-11824, 1994
A;Title: Primary structure of bovine collectin-43 (CL-43). Comparison with conglutinin a
A;Reference number: A53570; MUID:94216283; PMID:8163480
A;Accession: A53570
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-301 <LIM>
A;Cross-references: GB:X75912
R;Holmskov, U.; Teisner, B.; Willis, A.C.; Reid, K.B.; Jensenius, J.C.
J. Biol. Chem. 268, 10120-10125, 1993
A;Title: Purification and characterization of a bovine serum lectin (CL-43) with structu
A;Reference number: A46689; MUID:93252891; PMID:8486682
A;Accession: A46689
A;Molecule type: protein
A;Residues: 1-27 <HOL>
A;Experimental source: serum
A;Note: sequence extracted from NCBI backbone (NCBIP:131234)

```
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
C;Keywords: lectin
F;177-299/Domain: C-type lectin homology <LCH>

Alignment Scores:
Pred. No.:                3.36e-29        Length:          301
Score:                    418.00          Matches:         92
Percent Similarity:       46.18%          Conservative:    47
Best Local Similarity:    30.56%          Mismatches:      84
Query Match:              30.62%          Indels:          78
DB:                       2               Gaps:            9

US-10-054-536-2 (1-747) x A53570 (1-301)

Qy         55 TACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGT 114
              |||||||||           :::                     ::::::   |||
Db          6 TyrSerGluLysThrLeuThrAspProCys---------------ThrLeuValValCys 20

Qy        115 AGCTCTCCA------GGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAG 168
              :::   |||         :::   |||      |||:::||||||::::::|||   :::
Db         21 AlaProProAlaAspSerLeuArgGlyHisAspGlyArgAspGlyLysGluGlyProGln 40

Qy        169 GGAGAAAAGGGGGAACCAGGC-----------CAAGGGCTCAGAGGCTTACAGGGCCCC 216
              ||||||||||:::||||||               |||      |||   :::||||||
Db         41 GlyGluLysGlyAspProGlyProProGlyMetProGlyProAlaGlyArgGluGlyPro 60

Qy        217 CCTGGAAAG---------TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGA 267
                 |||:::         :::||||||||||   |||||||||   |||   ||||||
Db         61 SerGlyArgGlnGlySerMetGlyProProGlyThrProGlyProLysGlyGluProGly 80

Qy        268 CCA---------------------------------------------------------- 270
              |||
Db         81 ProGluGlyGlyValGlyAlaProGlyMetProGlySerProGlyProAlaGlyLeuLys 100

Qy        270 ---------------------------------------------------------- 270

Db        101 GlyGluArgGlyAlaProGlyProGlyGlyAlaIleGlyProGlnGlyProSerGlyAla 120

Qy        271 ------------------AAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGATGGT 309
                               ||||||   :::||||||||||||   |||         |||
Db        121 MetGlyProProGlyLeuLysGlyAspArgGlyAspProGlyGluLysGlyAlaArgGly 140

Qy        310 GATAGTAGCCTG-----------------------------------GCTGCCTCAGAA 333
              ::::::|||:::                                       |||
Db        141 GluThrSerValLeuGluValAspThrLeuArgGlnArgMetArgAsnLeuGluGlyGlu 160

Qy        334 AGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGC 393
              :::   ||||||      :::   :::   :::|||   :::   |||      |||
Db        161 ValGlnArgLeuGlnAsnIleValThrGlnTyrArgLysAlaValLeuPheProAspGly 180

Qy        394 AAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTG 453
              :::   ||||||   |||   |||   |||   |||   :::   ::::::
Db        181 GlnAlaValGlyGluLysIlePheLysThrAlaGlyAlaValLysSerTyrSerAspAla 200

Qy        454 AAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAAT 513
              :::   ||||||   :::   :::       :::|||:::||||||::::::|||||||||
Db        201 GluGlnLeuCysArgGluAlaLysGlyGlnLeuAlaSerProArgSerSerAlaGluAsn 220

Qy        514 GGAGCCATTCAGAATCTCATCAAG---------GAGGAAGCCTTCCTGGGCATCACTGAT 564
                 |||:::      |||::::::          :::   |||:::|||   :::   |||
Db        221 GluAlaValThrGlnLeuValArgAlaLysAsnLysHisAlaTyrLeuSerMetAsnAsp 240

Qy        565 GAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGG 624
                    ||||||:::|||         ||||||      |||   |||:::||||||
Db        241 IleSerLysGluGlyLysPheThrTyrProThrGlyGlySerLeuAspTyrSerAsnTrp 260

Qy        625 AACGAGGGTGAACCCAACAATGCTGGTTCTGAT---------GAAGATTGTGTATTGCTA 675
                    ||||||||||||||         |||         |||:::|||:::   :::
Db        261 AlaProGlyGluProAsnAsnArgAlaLysAspGluGlyProGluAsnCysLeuGluIle 280

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                 :::|||   |||||||||:::   |||             |||   :::||||||
Db        281 TyrSerAspGlyAsnTrpAsnAspIleGluCysArgGluGluArgLeuValIleCysGlu 300
```

```
Qy        736 TTC 738
                 |||
Db        301 Phe 301

RESULT 11
LNRBPS
pulmonary surfactant protein A precursor - rabbit
N;Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C;Accession: A29931
R;Boggaram, V.; Qing, K.; Mendelson, C.R.
J. Biol. Chem. 263, 2939-2947, 1988
A;Title: The major apoprotein of rabbit pulmonary surfactant. Elucidation of primary seq
A;Reference number: A29931; MUID:88139348; PMID:2830270
A;Accession: A29931
A;Molecule type: mRNA
A;Residues: 1-247 <BOG>
A;Cross-references: GB:J03542; NID:g165705; PIDN:AAA31465.1; PID:g165706
A;Note: 12-Pro was also found
A;Note: two species of mRNA, which appear to be transcribed from a single gene, could pr
A;Note: the amino end of the mature protein is blocked
C;Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t
C;Comment: This protein is a sialoglycoprotein synthesized by alveolar type II cells. It
pendent on the presence of calcium ions.
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; glyco
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-247/Product: pulmonary surfactant protein A #status predicted <MAT>
F;27-99/Region: collagen-like
F;126-245/Domain: C-type lectin homology <LCH>
F;16/Modified site: acetylated amino end (Ser) (in mature form) #status predicted
F;206/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.:                3.64e-29        Length:          247
Score:                    417.50          Matches:         91
Percent Similarity:       52.17%          Conservative:    41
Best Local Similarity:    35.97%          Mismatches:      94
Query Match:              30.59%          Indels:          27
DB:                       1               Gaps:            6

US-10-054-536-2 (1-747) x LNRBPS (1-247)

Qy         28 CTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCCAA 87
              |||   |||:::::::::::::|||      |||:::|||          :::|||
Db          6 LeuAlaLeuThrLeuIleSerAlaProAlaSerAspThrCysAspThrLysAspValCys 25

Qy         88 AAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAA 147
                     |||   :::            :::||||||   :::|||   |||||||:::
Db         26 IleGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuProGlyArg 43

Qy        148 GATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAA-------------- 192
              ||||||||||||||   ||||||:::   |||   ||||||
Db         44 AspGlyArgAspGlyValLysGlyAspProGlyProProGlyProMetGlyProProGly 63

Qy        193 GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCT 252
              |||:::   ||||||   |||      |||   :::|||   ||||||   ||||||
Db         64 GlyMetProGlyLeuProGlyArgAspGlyLeuIleGlyAlaProGlyValProGlyGlu 83

Qy        253 TCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGAT 312
                 |||      |||      |||:::::::|||   ||||||
Db         84 ArgGlyAspLysGlyGluProGlyGluArgGlyProProGly------------------ 97

Qy        313 AGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAG 372
                    |||   |||          :::   ||||||      :::      ::::::
Db         98 ------LeuProAlaTyrLeuAspGluGluLeuGlnAlaThrLeuHisGluLeuArgHis 115

Qy        373 TGGCTGACCTTCTCTCTGGGC----------------------AAACAAGTTGGGAAC 408
                       |||:::|||                          |||   ||||||
Db        116 HisAlaLeuGlnSerIleGlyValLeuSerLeuGlnGlySerMetLysAlaValGlyGlu 135
```

```
Qy        409 AAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTC 468
Db        136 LysIlePheSerThrAsnGlyGlnSerValAsnPheAspAlaIleArgGluValCysAla 155

Qy        469 AAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAAT 528
Db        156 ArgAlaGlyGlyArgIleAlaValProArgSerLeuGluGluAsnGluAlaIleAlaSer 175

Qy        529 CTCATCAAGGAGGAA---------GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGG 579
Db        176 IleValLysGluArgAsnThrTyrAlaTyrLeuGlyLeuAlaGluGlyProThrAlaGly 195

Qy        580 CAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCC 639
Db        196 AspPheTyrTyrLeuAspGlyAspProValAsnTyrThrAsnTrpTyrProGlyGluPro 215

Qy        640 AACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGAC 699
Db        216 ArgGlyGlnGlyArg---GluLysCysValGluMetTyrThrAspGlyLysTrpAsnAsp 234

Qy        700 GTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
Db        235 LysAsnCysLeuGlnTyrArgLeuValIleCysGluPhe 247
```

RESULT 12
S33603
surfactant protein D - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999
C;Accession: S33603
R;Lim, B.L.; Lu, J.; Reid, K.B.M.
Immunology 78, 159-165, 1993
A;Title: Structural similarity between bovine conglutinin and bovine lung surfactant pro
A;Reference number: S33603; MUID:93170856; PMID:8436402
A;Accession: S33603
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-369 <LIM>
C;Superfamily: pulmonary surfactant protein D; C-type lectin homology
F;248-367/Domain: C-type lectin homology <LCH>

Alignment Scores:

Pred. No.:	4.69e-29	Length:	369
Score:	416.50	Matches:	108
Percent Similarity:	39.46%	Conservative:	38
Best Local Similarity:	29.19%	Mismatches:	99
Query Match:	30.51%	Indels:	125
DB:	2	Gaps:	10

US-10-054-536-2 (1-747) x S33603 (1-369)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
Db          1 MetLeuLeuLeuProLeuSerValLeuLeuLeuLeuThrGlnProTrpArgSerLeuGly 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCT---------GCAGTGATTGCC 111
Db         21 AlaGluMetLysIleTyr---SerGlnLysThrMetAlaAsnAlaCysThrLeuValMet 39

Qy        112 TGTAGCTCTCCA------GGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACC 165
Db         40 CysSerProProGluAspGlyLeuProGlyArgAspGlyArgAspGlyArgGluGlyPro 59

Qy        166 AAGGGAGAAAAGGGGGAACCAGGCCAA------------------------------ 192
Db         60 ArgGlyGluLysGlyAspProGlySerProGlyProAlaGlyArgAlaGlyMetProGly 79

Qy        193 ----------------------------------------------------GGGCTCAGA 201
Db         80 ProAlaGlyProIleGlyLeuLysGlyAspAsnGlySerAlaGlyGluProGlyProLys 99

Qy        202 GGCTTACAGGGCCCCCCT---------------------------------------- 219
```

```
Db        100 GlyAspThrGlyProProGlyProProGlyMetProGlyProAlaGlyArgGluGlyPro 119

Qy        220 ------------GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCT--------------- 252
Db        120 SerGlyLysGlnGlySerMetGlyProProGlyThrProGlyProLysGlyAspThrGly 139

Qy        253 -----------------------------TCTGGGTCACCAGGACCAAAGGGCCAAAAA 282
Db        140 ProLysGlyGlyValGlyAlaProGlyIleGlnGlySerProGlyProAlaGlyLeuLys 159

Qy        283 ------------------GGAGACCCTGGA---------------------------- 294
Db        160 GlyGluArgGlyAlaProGlyAspProGlyAlaProGlyArgAlaGlyAlaProGlyPro 179

Qy        294 ------------------------------------------------------------ 294
Db        180 ArgGlyAlaIleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGly 199

Qy        295 ---------------------AAAAGTCCGGATGGTGATAGTAGCCTGGCT--------- 324
Db        200 AspArgGlyThrProGlyGluArgGlyAlaLysGlyGluSerGlyLeuAlaGluValAsn 219

Qy        325 --------------------------GCCTCAGAAAGAAAAGCTCTGCAAACAGAAATG 357
Db        220 AlaLeuArgGlnArgValGlyIleLeuGluGlyGlnLeuGlnArgLeuGlnAsnAlaPhe 239

Qy        358 GCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTC 417
Db        240 SerGlnTyrLysLysAlaMetLeuPheProAsnGlyArgSerValGlyGluLysIlePhe 259

Qy        418 CTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAG 477
Db        260 LysThrValGlySerGluLysThrPheGlnAspAlaGlnGlnIleCysThrGlnAlaGly 279

Qy        478 GCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC--- 534
Db        280 GlyGlnLeuProSerProArgSerGlyAlaGluAsnGluAlaLeuThrGlnLeuAlaThr 299

Qy        535 ------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTG 588
Db        300 AlaGlnAsnLysAlaAlaPheLeuSerMetSerAspThrArgLysGluGlyThrPheIle 319

Qy        589 GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT 648
Db        320 TyrProThrGlyGluProLeuValTyrSerAsnTrpAlaProGlnGluProAsnAsnAsp 339

Qy        649 GGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGC 708
Db        340 GlyGlySerGluAsnCysValGluIleProAsnGlyLysTrpAsnAspLysValCys 359

Qy        709 TCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
Db        360 GlyGluGlnArgLeuValIleCysGluPhe 369
```

RESULT 13
LNRTPS
pulmonary surfactant protein A precursor - rat
N;Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C;Accession: A29299; JS0034; S23183
R;Sano, K.; Fisher, J.; Mason, R.J.; Kuroki, Y.; Schilling, J.; Benson, B.; Voelker, D.
Biochem. Biophys. Res. Commun. 144, 367-374, 1987
A;Title: Isolation and sequence of a cDNA clone for the rat pulmonary surfactant-associ
A;Reference number: A29299; MUID:87213191; PMID:3579914
A;Contents: 0.9 kb cDNA
A;Accession: A29299
A;Molecule type: mRNA
A;Residues: 1-248 <SAN>
A;Note: part of the sequence, including the amino end of the mature protein, was confir
R;Fisher, J.H.; Emrie, P.A.; Shannon, J.; Sano, K.; Hattler, B.; Mason, R.J.
Biochim. Biophys. Acta 950, 338-345, 1988
A;Title: Rat pulmonary surfactant protein A is expressed as two differently sized mRNA

```
A;Reference number: JS0034; MUID:89000785; PMID:2901856
A;Contents: 1.6 kb cDNA
A;Accession: JS0034
A;Molecule type: mRNA
A;Residues: 1-248 <FIS>
A;Note: the codons given for residues 78, 84, and 180 are inconsistent with the authors'
R;Lacaze-Masmonteil, T.; Fraslon, C.; Bourbon, J.; Raymondjean, M.; Kahn, A.
Eur. J. Biochem. 206, 613-623, 1992
A;Title: Characterization of the rat pulmonary surfactant protein A promoter.
A;Reference number: S23183; MUID:92298987; PMID:1606951
A;Accession: S23183
A;Molecule type: DNA
A;Residues: 1-32 <LAC>
C;Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t
C;Comment: This protein is a sialoglycoprotein synthesized by alveolar type II cells. It
pendent on the presence of calcium ions.
C;Comment: Two species of mRNA, which probably are transcribed from a single gene, have
C;Comment: Size heterogeneity of these proteins arises from posttranslational modificati
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: alveolar proteinosis; calcium; gaseous exchange; glycoprotein; hydroxyprolin
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-248/Product: pulmonary surfactant protein A #status experimental <MAT>
F;37-109/Region: collagen-like
F;127-246/Domain: C-type lectin homology <LCH>
F;21/Binding site: carbohydrate (Asn) (covalent) #status absent
F;30,33,36,42,54,57,63,67,70,76/Modified site: 4-hydroxyproline (Pro) #status experiment
F;207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.:                5.86e-28      Length:          248
Score:                    404.00        Matches:         99
Percent Similarity:       51.34%        Conservative:    35
Best Local Similarity:    37.93%        Mismatches:      99
Query Match:              29.60%        Indels:          28
DB:                       1             Gaps:            9

US-10-054-536-2 (1-747) x LNRTPS (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||||||||      ||||||          |||          |||   |||
Db          1 MetSerLeuCys---SerLeuAlaPheThrLeuPheLeuThrValValAlaGlyIleLys 19

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
                 ||||||      |||              |||   :::         :::
Db         20 CysAsnValThr------AspValCysAlaGlySerProGlyIle------ProGlyAla 35

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||   :::|||   ||||||:::||||||||||||||||   ||||||:::   |||
Db         36 ProGlyAsnHisGlyLeuProGlyArgAspGlyArgAspGlyValLysGlyAspProGly 55

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTG------------ 228
                ||||||                ||||||||||||   :::
Db         56 ProProGlyPro--------------MetGlyProProGlyGlyMetProGlyLeuPro 70

Qy        229 GGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGAC 288
              |||      |||  ||||||      |||:::||||||   :::|||   ||||||:::
Db         71 GlyArgAspGlyLeuProGlyAlaProGlyAlaProGlyGluArgGlyAspLysGlyGlu 90

Qy        289 CCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAA 348
              ||||||:::         |||              |||          :::   ||||||
Db         91 ProGlyGluArg------GlyLeuProGlyPheProAlaTyrLeuAspGluGluLeuGln 108

Qy        349 ACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGC--------------- 393
              ||||||:::      ||||||       :::      :::::::|||
Db        109 ThrGluLeuTyrGluIleLysHisGlnIleLeuGlnThrMetGlyValLeuSerLeuGln 128

Qy        394 ---------AAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                      ||||||:::|||   |||   |||||||||:::   :::   |||
Db        129 GlySerMetLeuSerValGlyAspLysValPheSerThrAsnGlyGlnSerValAsnPhe 148

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
              :::   :::|||   :::|||   :::          :::::::|||   ||||||
Db        149 AspThrIleLysGluMetCysThrArgAlaGlyGlyAsnIleAlaValProArgThrPro 168

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
                ||||||   ||||||   ::::::   |||:::                :::||||||
Db        169 GluGluAsnGluAlaIleAlaSerIleAlaLysLysTyrAsnAsnTyrValTyrLeuGly 188

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
              :::   :::::::::|||   |||   |||       |||   |||       :::   |||
Db        189 MetIleGluAspGlnThrProGlyAspPheHisTyrLeuAspGlyAlaSerValAsnTyr 208

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
              |||||||||      |||||||||         |||       |||   ||||||   :::
Db        209 ThrAsnTrpTyrProGlyGluProArgGlyGlnGlyLys---GluLysCysValGluMet 227

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                  :::|||   |||||||||         |||                ||||||||||||||
Db        228 TyrThrAspGlyThrTrpAsnAspArgGlyCysLeuGlnTyrArgLeuAlaValCysGlu 247

Qy        736 TTC 738
              |||
Db        248 Phe 248

RESULT 14
LNHUP1
pulmonary surfactant protein A precursor (clone 1A) - human
N;Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C;Accession: B25720
R;Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman, L.;
J. Biol. Chem. 261, 9029-9033, 1986
A;Title: Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfacta
A;Reference number: A25720; MUID:86250832; PMID:3755136
A;Accession: B25720
A;Molecule type: mRNA
A;Residues: 1-248 <FLO>
A;Cross-references: GB:K03475
A;Note: part of the sequence was confirmed by protein sequencing
A;Note: the amino end of the mature protein, which was not identified, is partially acet
A;Note: clones corresponding to two different proteins were sequenced. Cotranslational m
C;Genetics:
A;Gene: GDB:SFTPA1; SFTP1; SP-A; SP-A1
A;Cross-references: GDB:119593; OMIM:178630
A;Map position: 10q22-10q23
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; glycc
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F;127-246/Domain: C-type lectin homology <LCH>
F;21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted
F;30,33,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status pred
F;207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.:                2.23e-27      Length:          248
Score:                    397.50        Matches:         96
Percent Similarity:       50.57%        Conservative:    36
Best Local Similarity:    36.78%        Mismatches:      90
Query Match:              29.12%        Indels:          39
DB:                       1             Gaps:            11

US-10-054-536-2 (1-747) x LNHUP1 (1-248)

Qy         22 CCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGAT 81
              ||||||   |||      ::::::   |||:::                  :::|||
Db          5 ProLeuAlaLeuThrLeuIleLeuMetAlaAlaSerGlyAlaAlaCysGluValLysAsp 24

Qy         82 GCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCA 141
                           |||   :::            :::||||||   :::|||   |||
Db         25 ValCysValGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuPro 42

Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
              |||:::|||||||||||||||   ||||||:::   |||   ||||||
Db         43 GlyArgAspGlyArgAspGlyLeuLysGlyAspProGlyProProGlyPro--------- 59
```

```
Qy        202 GGCTTACAGGGCCCCCCCTGGAAAGTTGGGG---CCTCCAGGAAAT---------CCAGGG 249
                ||||||||||||:::          ||||||||||||          ||||||
Db         60 ------MetGlyProProGlyGluThrProCysProProGlyAsnAsnGlyLeuProGly 77

Qy        250 CCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGAT 306
                |||   ||||||   :::|||:::||||||:::||||||   :::   |||
Db         78 AlaProGlyValProGlyGluArgGlyGluLysGlyGluProGlyGluArgGlyProPro 97

Qy        307 GGT---------------GATAGTAGCCTGGCTGCCTCA------------GAAAGAAAA 339
                |||               |||      |||   |||:::                :::
Db         98 GlyLeuProAlaHisLeuAspGluGluLeuGlnAlaThrLeuHisAspPheArgHisGln 117

Qy        340 GCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAA 399
                |||||||||      |||                                |||
Db        118 IleLeuGlnThrArgGlyAla-------------------------LeuSerLeuGln 128

Qy        400 ---------------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                              ||||||   |||   |||   :::||||||:::   :::||||||
Db        129 GlySerIleMetThrValGlyGluLysValPheSerSerAsnGlyGlnSerIleThrPhe 148

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
                :::   ::::::      |||   :::           :::|||   |||||||||
Db        149 AspAlaIleGlnGluAlaCysAlaArgAlaGlyGlyArgIleAlaValProArgAsnPro 168

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
                ||||||   ||||||   :::   :::|||:::             |||::::::|||
Db        169 GluGluAsnGluAlaIleAlaSerPheValLysLysTyrAsnThrTyrAlaTyrValGly 188

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
                :::|||:::      :::   |||   |||           |||      :::   |||
Db        189 LeuThrGluGlyProSerProGlyAspPheArgTyrSerAspGlyThrProValAsnTyr 208

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
                |||||||||      |||||||||          |||      |||   ||||||   :::
Db        209 ThrAsnTrpTyrArgGlyGluProAlaGlyArgGlyLys---GluGlnCysValGluMet 227

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                    :::||||||||||||||||      |||      |||   |||   :::||||||
Db        228 TyrThrAspGlyGlnTrpAsnAspArgAsnCysLeuTyrSerArgLeuThrIleCysGlu 247

Qy        736 TTC 738
                |||
Db        248 Phe 248

RESULT 15
LNHUP6
pulmonary surfactant protein A precursor (clone 6A) - human
N;Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C;Accession: A25720
R;Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman, L.;
J. Biol. Chem. 261, 9029-9033, 1986
A;Title: Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfacta
A;Reference number: A25720; MUID:86250832; PMID:3755136
A;Accession: A25720
A;Molecule type: mRNA
A;Residues: 1-248 <FLO>
A;Cross-references: GB:M13686; NID:g190669; PIDN:AAA60211.1; PID:g190670
A;Note: part of the sequence was confirmed by protein sequencing
A;Note: the amino end of the mature protein, which was not identified, is partially acet
A;Note: clones corresponding to two different proteins were sequenced. Cotranslational m
C;Genetics:
A;Gene: GDB:SFTPA1; SFTP1; SP-A; SP-A1
A;Cross-references: GDB:119593; OMIM:178630
A;Map position: 10q22-10q23
C;Superfamily: mannose-binding lectin; C-type lectin homology
C;Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; glycc
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F;127-246/Domain: C-type lectin homology <LCH>
F;21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted
F;30,33,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status pre
F;207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.:                5.09e-27        Length:          248
Score:                    393.50          Matches:         96
Percent Similarity:       50.57%          Conservative:    36
Best Local Similarity:    36.78%          Mismatches:      90
Query Match:              28.83%          Indels:          39
DB:                       1               Gaps:            11

US-10-054-536-2 (1-747) x LNHUP6 (1-248)

Qy         22 CCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGAT 81
                ||||||   |||      ::::::   |||:::                  :::|||
Db          5 ProLeuAlaLeuAsnLeuIleLeuMetAlaAlaSerGlyAlaAlaCysGluValLysAsp 24

Qy         82 GCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCA 141
                           |||   :::           :::||||||   :::|||   |||
Db         25 ValCysValGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuPro 42

Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
                |||:::|||||||||||||||   ||||||:::   |||   ||||||
Db         43 GlyArgAspGlyArgAspGlyLeuLysGlyAspProGlyProProGlyPro--------- 59

Qy        202 GGCTTACAGGGCCCCCCCTGGAAAGTTGGGG---CCTCCAGGAAAT---------CCAGGG 249
                ||||||||||||:::::      ||||||||||||          ||||||
Db         60 ------MetGlyProProGlyGluMetProCysProProGlyAsnAspGlyLeuProGly 77

Qy        250 CCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGAT 306
                |||   ||||||      |||:::||||||:::||||||   :::   |||
Db         78 AlaProGlyIleProGlyGluCysGlyGluLysGlyGluProGlyGluArgGlyProPro 97

Qy        307 GGT---------------GATAGTAGCCTGGCTGCCTCA------------GAAAGAAAA 339
                |||               |||      |||   |||:::                :::
Db         98 GlyLeuProAlaHisLeuAspGluGluLeuGlnAlaThrLeuHisAspPheArgHisGln 117

Qy        340 GCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAA 399
                |||||||||      |||                                |||
Db        118 IleLeuGlnThrArgGlyAla-------------------------LeuSerLeuGln 128

Qy        400 ---------------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                              ||||||   |||   |||   :::||||||:::   :::||||||
Db        129 GlySerIleMetThrValGlyGluLysValPheSerSerAsnGlyGlnSerIleThrPhe 148

Qy        445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
                :::   ::::::      |||   :::           :::|||   |||||||||
Db        149 AspAlaIleGlnGluAlaCysAlaArgAlaGlyGlyArgIleAlaValProArgAsnPro 168

Qy        505 GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGC 555
                ||||||   ||||||   :::   :::|||:::             |||::::::|||
Db        169 GluGluAsnGluAlaIleAlaSerPheValLysLysTyrAsnThrTyrAlaTyrValGly 188

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
                :::|||:::      :::   |||   |||           |||      :::   |||
Db        189 LeuThrGluGlyProSerProGlyAspPheArgTyrSerAspGlyThrProValAsnTyr 208

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
                |||||||||      |||||||||          |||      |||   ||||||   :::
Db        209 ThrAsnTrpTyrArgGlyGluProAlaGlyArgGlyLys---GluGlnCysValGluMet 227

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                    :::||||||||||||||||      |||      |||   |||   :::||||||
Db        228 TyrThrAspGlyGlnTrpAsnAspArgAsnCysLeuTyrSerArgLeuThrIleCysGlu 247

Qy        736 TTC 738
                |||
Db        248 Phe 248

Search completed: June 18, 2004, 21:03:40
Job time : 26 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 21, 2004, 17:18:48 ; Search time 81 Seconds
(without alignments)
5117.880 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 747
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	747	100.0	900	4	US-09-198-603C-1	Sequence 1, Appli
2	204.4	27.4	1211	4	US-09-198-603C-25	Sequence 25, Appl
3	126	16.9	714	4	US-09-198-603C-26	Sequence 26, Appl
4	73.4	9.8	1558	4	US-09-198-603C-24	Sequence 24, Appl
5	67.6	9.0	885	1	US-08-365-103B-3	Sequence 3, Appli
6	67.6	9.0	924	1	US-08-365-103B-5	Sequence 5, Appli
7	67.6	9.0	1005	1	US-08-365-103B-1	Sequence 1, Appli
8	62.8	8.4	369	4	US-09-535-521-24	Sequence 24, Appl
C 9	62.8	8.4	369	4	US-09-535-521-26	Sequence 26, Appl
10	62.8	8.4	384	4	US-09-535-521-7	Sequence 7, Appli
C 11	62.8	8.4	384	4	US-09-535-521-9	Sequence 9, Appli
12	62.8	8.4	417	4	US-09-535-521-10	Sequence 10, Appl
C 13	62.8	8.4	417	4	US-09-535-521-12	Sequence 12, Appl
14	62.8	8.4	423	4	US-09-535-521-13	Sequence 13, Appl
C 15	62.8	8.4	423	4	US-09-535-521-15	Sequence 15, Appl
16	62.8	8.4	561	4	US-09-535-521-16	Sequence 16, Appl
C 17	62.8	8.4	561	4	US-09-535-521-18	Sequence 18, Appl
18	62.8	8.4	624	4	US-09-535-521-19	Sequence 19, Appl
C 19	62.8	8.4	624	4	US-09-535-521-21	Sequence 21, Appl
20	62.8	8.4	876	4	US-09-535-521-4	Sequence 4, Appli
C 21	62.8	8.4	876	4	US-09-535-521-6	Sequence 6, Appli
22	62.8	8.4	2851	4	US-09-535-521-1	Sequence 1, Appli
C 23	62.8	8.4	2851	4	US-09-535-521-3	Sequence 3, Appli
C 24	55.2	7.4	7218	1	US-08-232-463-14	Sequence 14, Appl
25	52.4	7.0	1152	4	US-09-776-976-1	Sequence 1, Appli
26	52.4	7.0	1152	4	US-09-909-547-1	Sequence 1, Appli
27	51.4	6.9	1276	2	US-08-463-911-1	Sequence 1, Appli
28	51.4	6.9	1276	4	US-09-776-976-3	Sequence 3, Appli
29	51.4	6.9	1276	4	US-09-909-547-3	Sequence 3, Appli
30	50.2	6.7	6674	4	US-09-620-312D-110	Sequence 110, App
31	48.6	6.5	1313	2	US-08-463-911-6	Sequence 6, Appli
32	48.6	6.5	4517	3	US-09-140-804-9	Sequence 9, Appli
33	48.6	6.5	4517	4	US-09-686-838B-9	Sequence 9, Appli
34	48.6	6.5	4517	4	US-09-776-976-5	Sequence 5, Appli
35	48.6	6.5	4517	4	US-09-909-547-5	Sequence 5, Appli
36	48.6	6.5	4545	4	US-09-569-852B-5	Sequence 5, Appli
37	47.2	6.3	855	4	US-09-552-204A-10	Sequence 10, Appl
38	47	6.3	3389	4	US-09-620-312D-1061	Sequence 1061, Ap
39	46	6.2	511	4	US-09-702-705-1138	Sequence 1138, Ap
40	46	6.2	511	4	US-09-736-457-1138	Sequence 1138, Ap
41	46	6.2	511	4	US-09-614-124B-1138	Sequence 1138, Ap
42	46	6.2	511	4	US-09-671-325-1138	Sequence 1138, Ap
43	45.2	6.1	729	3	US-09-140-804-10	Sequence 10, Appl
44	45.2	6.1	729	4	US-09-686-838B-10	Sequence 10, Appl
45	45.2	6.1	2487	4	US-09-620-312D-160	Sequence 160, App

ALIGNMENTS

RESULT 1
US-09-198-603C-1
; Sequence 1, Application US/09198603C
; Patent No. 6337193
; GENERAL INFORMATION:
; APPLICANT: TULLY, Raymond E.
; APPLICANT: CALTAGIRONE, G. Thomas
; APPLICANT: MOYER, Shawn S.
; APPLICANT: RONNING, Michael T.
; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC
; TITLE OF INVENTION: YEAST
; FILE REFERENCE: A7290
; CURRENT APPLICATION NUMBER: US/09/198,603C
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Human
US-09-198-603C-1

Query Match 100.0%; Score 747; DB 4; Length 900;
Best Local Similarity 100.0%; Pred. No. 4.1e-228;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425
```

```
Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              ||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812
```

RESULT 2
US-09-198-603C-25
; Sequence 25, Application US/09198603C
; Patent No. 6337193
; GENERAL INFORMATION:
; APPLICANT: TULLY, Raymond E.
; APPLICANT: CALTAGIRONE, G. Thomas
; APPLICANT: MOYER, Shawn S.
; APPLICANT: RONNING, Michael T.
; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC
; TITLE OF INVENTION: YEAST
; FILE REFERENCE: A7290
; CURRENT APPLICATION NUMBER: US/09/198,603C
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 1211
; TYPE: DNA
; ORGANISM: Mouse
US-09-198-603C-25

Query Match 27.4%; Score 204.4; DB 4; Length 1211;
Best Local Similarity 71.7%; Pred. No. 5.1e-55;
Matches 268; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```
Qy        374 GGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAA 433
              || ||  |||||||||  | || |||||||  || ||||  ||  ||| || || | |||
Db        576 GGGTGCTCTTCTCTCTGAGTGAAAAAGTTGGAAAGAAGTATTTTGTGAGCAGTGTTAAAA 635

Qy        434 TAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCC 493
              ||||  |  ||||  | ||||||||||  ||||    | |  |||||||  ||||||||||| |
Db        636 AGATGAGCCTTGATAGAGTGAAGGCCCTGTGCTCCGAATTCCAGGGCTCTGTGGCCACTC 695

Qy        494 CCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGG 553
              ||||||||||||  || ||     |||||  |||||   |   |||  ||     ||||  | |||
Db        696 CCAGGAATGCTGAGGAAAACTCGGCCATCCAGAAAGTGGCCAAAGATATTGCCTACTTGG 755

Qy        554 GCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCT 613
              |||||||  ||||  ||  |   |||||     ||||  ||||||||||||||||  |||  ||   ||
Db        756 GCATCACAGATGTGAGGGTTGAAGGCAGTTTTGAGGATCTGACAGGAAACAGAGTGCGCT 815

Qy        614 ACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGC 673
              |  ||  ||  |||||  ||  ||  ||  ||||||||   |  ||    ||   |||||  |||||   ||
```

```
Db        816 ATACTAATTGGAATGATGGGGAGCCCAACAACACGGGCGATGGGGAAGACTGTGTGGTGA 875

Qy        674 TACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTG 733
              |  ||  |||||||  |||||||  ||  ||||||||||||   |||   | ||||  | ||||
Db        876 TCTTGGGAAATGGCAAGTGGAACGATGTCCCCTGCTCTGACTCTTTTTTGGCAATATGTG 935

Qy        734 AGTTCCCTATCTGA 747
              |  |||  ||  ||||
Db        936 AATTCTCTGACTGA 949
```

RESULT 3
US-09-198-603C-26
; Sequence 26, Application US/09198603C
; Patent No. 6337193
; GENERAL INFORMATION:
; APPLICANT: TULLY, Raymond E.
; APPLICANT: CALTAGIRONE, G. Thomas
; APPLICANT: MOYER, Shawn S.
; APPLICANT: RONNING, Michael T.
; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC
; TITLE OF INVENTION: YEAST
; FILE REFERENCE: A7290
; CURRENT APPLICATION NUMBER: US/09/198,603C
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 714
; TYPE: DNA
; ORGANISM: CHICKEN
US-09-198-603C-26

Query Match 16.9%; Score 126; DB 4; Length 714;
Best Local Similarity 53.0%; Pred. No. 3.8e-30;
Matches 351; Conservative 0; Mismatches 290; Indels 21; Gaps 3;

```
Qy         79 GATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTC 138
              ||  |   |||  |   |  |  |  | |||   |||||  ||||  |   ||||  ||  ||
Db         37 GAAGAGAAAATGTATTCCTGTCCCATCATTCAGTGTAGTGCTCCTGCAGTCAATGGATTA 96

Qy        139 CCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTC 198
              |||||||  ||||||  |  |||||  ||||  ||  ||||||||  ||  |||||  ||||  ||
Db         97 CCAGGCAGAGATGGAAGAGATGGTCCCAAAGGGGAAAAGGGAGACCCAGGAGAAGGACTG 156

Qy        199 AGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGG 258
              |||||  |  |||||    ||||||||     ||  ||  |  ||||      |||          ||
Db        157 AGAGGTCTGCAGGGTTTGCCTGGAAAAGCAGGACCCCAAGGATTAAAAGGAGAGGTGGGA 216

Qy        259 TCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAA----------AAAGTCCGGAT 306
              |||  ||||    ||  ||  ||||||||||| |  |||||                |   | | |
Db        217 CCACAAGGAGAGAAAGGTCAAAAAGGAGAACGTGGAATTGTTGTAACTGATGACCTGCAC 276

Qy        307 GGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
              |  | |  |        |   ||  |||     |   ||  ||   ||  |    |   |
Db        277 CGACAAATAACTGATTTGGAAGCAAAAATCCGGGTATTGGAAGATGACTTAAGCAGATAC 336

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              |||||      |||  ||        | |  | |  |||| || ||  | ||  |  | | |
Db        337 AAAAAAGCCTTGAGTTTAAAGGACGTCGTAAACATTGGTAAAAAAATGTTTGTCTCAACT 396

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              ||  |  |  |   |  |||||||| |  ||  ||  | ||||  ||           |
Db        397 GGAAAGAAATATAATTTTGAAAAGGGAAAATCCCTTTGTGCAAAAGCTGGAAGTGTGCTT 456

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG------GAG 540
              |||  |  ||  |||||  |   ||  ||||||   |||   |   |   |  |||  |
Db        457 GCCTCTCCTAGGAACGAGGCTGAGAATACAGCTTTAAAAGACTTAATTGACCCTTCAAGC 516

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||  |   |  ||  ||   ||||||    |  ||  ||  ||     ||   ||  | ||||   ||
Db        517 CAAGCTTATATTGGGATATCTGATGCACAAACTGAGGGCAGATTCATGTACCTGAGTGGT 576
```

```
Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              | || ||||  ||||||||    || ||||| || |||          |||||
Db        577 GGGCCTTTAACTTACAGCAACTGGAAACCTGGAGAACCAAATAATCACAAAA---ATGAA 633

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              || ||||   || ||    |   |||  | ||||||| |   ||| || |  ||  ||
Db        634 GACTGTGCGGTGATAGAAGACTCTGGAAAATGGAATGATTTAGACTGTTCAAATTCAAAT 693

Qy        721 CT 722
              |
Db        694 AT 695


RESULT 4
US-09-198-603C-24
; Sequence 24, Application US/09198603C
; Patent No. 6337193
; GENERAL INFORMATION:
;  APPLICANT:  TULLY, Raymond E.
;  APPLICANT:  CALTAGIRONE, G. Thomas
;  APPLICANT:  MOYER, Shawn S.
;  APPLICANT:  RONNING, Michael T.
;  TITLE OF INVENTION:  EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC
;  TITLE OF INVENTION:  YEAST
;  FILE REFERENCE:  A7290
;  CURRENT APPLICATION NUMBER: US/09/198,603C
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS:  26
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 1558
;   TYPE: DNA
;   ORGANISM: RAT
US-09-198-603C-24

  Query Match             9.8%;  Score 73.4;  DB 4;  Length 1558;
  Best Local Similarity   77.4%;  Pred. No. 3.5e-13;
  Matches   89;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         185 CAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATC 244
               |||| |||||||||| ||||| |||||||| || || ||  ||||||||||||||| |
Db        1173 CAGGTCAAGGGCTCAGGGGCTTGCAGGGCCCTCCAGGGAAACTGGGGCCTCCAGGAAGTG 1232

Qy         245 CAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAG 299
                |||  |  ||||    | ||||||||| |||||||||| || | |||| | ||
Db        1233 TAGGAGCCCCTGGAAGTCAAGGACCAAAAGGCCAAAAAGGGGATCGTGGAGACAG 1287


RESULT 5
US-08-365-103B-3
; Sequence 3, Application US/08365103B
; Patent No. 5766943
;  GENERAL INFORMATION:
;    APPLICANT:  Lynch, Richard G
;    APPLICANT:  Nunez, Raphael D.
;    APPLICANT:  Yodoi, Jungi
;    TITLE OF INVENTION:  DNA Sequences for Soluble Froms of CD23
;    TITLE OF INVENTION:  and Methods of Use for Same
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zarley, McKee, Thomte, Voorhees & Sease
;      STREET:  801 Grand Ave. Suite 3200
;      CITY:  Des Moines
;      STATE:  Iowa
;      COUNTRY:  United States
;      ZIP:  50309
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,103B
;      FILING DATE:  28-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nebel, Heidi S.
;      REGISTRATION NUMBER:  37,719
;      REFERENCE/DOCKET NUMBER:  Uirf N5-24
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (515) 288-3667
;      TELEFAX:  (515) 288-1338
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  24..884
US-08-365-103B-3

  Query Match             9.0%;  Score 67.6;  DB 1;  Length 885;
  Best Local Similarity   55.6%;  Pred. No. 1.8e-11;
  Matches  130;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy         484 GTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAA 543
               ||   || || |||  |    | | || |       || ||  | ||| |||| ||||
Db         567 GTCAGCATCCACAGCCAAAAGGAACAGGACTTCCTGATGCAACACATCAACAAGAAGGAT 626

Qy         544 GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAAT 603
               |||   | ||| ||   |||   || |  || ||  |||||||      |   || |
Db         627 TCCTGGATTGGCCTCCAGGATCTCAATATGGAGGGAGAGTTTGTATGGTCGGACGGGAGC 686

Qy         604 AGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGAT 663
                 ||   || |  ||||||||    || || ||||| || |  ||    | ||| ||
Db         687 CCTGTGGGTTATAGCAACTGGAATCCAGGGGAGCCCAATAACGGGGGCCAGGGTGAGGAC 746

Qy         664 TGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
               |||||  || | | |  |  |||||||||| |||| |  |||| || || |
Db         747 TGTGTGATGATGCGGGGATCCGGCCAGTGGAACGACGCCTTCTGCCGCAGCTAC 800


RESULT 6
US-08-365-103B-5
; Sequence 5, Application US/08365103B
; Patent No. 5766943
;  GENERAL INFORMATION:
;    APPLICANT:  Lynch, Richard G
;    APPLICANT:  Nunez, Raphael D.
;    APPLICANT:  Yodoi, Jungi
;    TITLE OF INVENTION:  DNA Sequences for Soluble Froms of CD23
;    TITLE OF INVENTION:  and Methods of Use for Same
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zarley, McKee, Thomte, Voorhees & Sease
;      STREET:  801 Grand Ave. Suite 3200
;      CITY:  Des Moines
;      STATE:  Iowa
;      COUNTRY:  United States
;      ZIP:  50309
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,103B
;      FILING DATE:  28-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nebel, Heidi S.
```

```
;       REGISTRATION NUMBER:  37,719
;       REFERENCE/DOCKET NUMBER:  Uirf N5-24
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (515) 288-3667
;       TELEFAX:  (515) 288-1338
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  924 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  24..923
US-08-365-103B-5

  Query Match             9.0%;  Score 67.6;  DB 1;  Length 924;
  Best Local Similarity   55.6%;  Pred. No. 1.8e-11;
  Matches  130;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        484 GTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAA 543
              ||   || || |||  |    | | || |       || ||  |  ||| |||| ||||
Db        606 GTCAGCATCCACAGCCAAAAGGAACAGGACTTCCTGATGCAACACATCAACAAGAAGGAT 665

Qy        544 GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAAT 603
              |||   | |||  ||   |||   || |  ||  ||  |||||||      |   ||| |
Db        666 TCCTGGATTGGCCTCCAGGATCTCAATATGGAGGGAGAGTTTGTATGGTCGGACGGGAGC 725

Qy        604 AGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGAT 663
              ||   || |  ||||||||    || || ||||| || |  ||    | ||| ||
Db        726 CCTGTGGGTTATAGCAACTGGAATCCAGGGGAGCCCAATAACGGGGGCCAGGGTGAGGAC 785

Qy        664 TGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              |||||  || | | |  |   |||||||||| |||| |  ||||  || || |
Db        786 TGTGTGATGATGCGGGGATCCGGCCAGTGGAACGACGCCTTCTGCCGCAGCTAC 839


RESULT 7
US-08-365-103B-1
; Sequence 1, Application US/08365103B
; Patent No. 5766943
;  GENERAL INFORMATION:
;    APPLICANT:  Lynch, Richard G
;    APPLICANT:  Nunez, Raphael D.
;    APPLICANT:  Yodoi, Jungi
;    TITLE OF INVENTION:  DNA Sequences for Soluble Froms of CD23
;    TITLE OF INVENTION:  and Methods of Use for Same
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zarley, McKee, Thomte, Voorhees & Sease
;      STREET:  801 Grand Ave. Suite 3200
;      CITY:  Des Moines
;      STATE:  Iowa
;      COUNTRY:  United States
;      ZIP:  50309
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,103B
;      FILING DATE:  28-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nebel, Heidi S.
;      REGISTRATION NUMBER:  37,719
;      REFERENCE/DOCKET NUMBER:  Uirf N5-24
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (515) 288-3667
;      TELEFAX:  (515) 288-1338
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1005 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  24..1004
US-08-365-103B-1

  Query Match             9.0%;  Score 67.6;  DB 1;  Length 1005;
  Best Local Similarity   55.6%;  Pred. No. 1.9e-11;
  Matches  130;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        484 GTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAA 543
              ||   || || |||  |    | | || |       || ||  |  ||| |||| ||||
Db        687 GTCAGCATCCACAGCCAAAAGGAACAGGACTTCCTGATGCAACACATCAACAAGAAGGAT 746

Qy        544 GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAAT 603
              |||   | |||  ||   |||   || |  ||  ||  |||||||      |   ||| |
Db        747 TCCTGGATTGGCCTCCAGGATCTCAATATGGAGGGAGAGTTTGTATGGTCGGACGGGAGC 806

Qy        604 AGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGAT 663
              ||   || |  ||||||||    || || ||||| || |  ||    | ||| ||
Db        807 CCTGTGGGTTATAGCAACTGGAATCCAGGGGAGCCCAATAACGGGGGCCAGGGTGAGGAC 866

Qy        664 TGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              |||||  || | | |  |   |||||||||| |||| |  ||||  || || |
Db        867 TGTGTGATGATGCGGGGATCCGGCCAGTGGAACGACGCCTTCTGCCGCAGCTAC 920


RESULT 8
US-09-535-521-24
; Sequence 24, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 369
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(369)
US-09-535-521-24

  Query Match             8.4%;  Score 62.8;  DB 4;  Length 369;
  Best Local Similarity   57.7%;  Pred. No. 3.7e-10;
  Matches  112;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||   || |||| |||    |||   | ||| ||   ||    | | | ||| ||| ||||
Db        165 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 224

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |      ||   | ||     |||| ||| |  ||||||       || || |||||||||
Db        225 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 284
```

```
Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
              |  ||     |  || || || ||  || | | |      || ||||||||||||| |
Db        285 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 344

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db        345 CTGCGGCAGCTCGC 358


RESULT 9
US-09-535-521-26/c
; Sequence 26, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 369
;   TYPE: DNA
;   ORGANISM: Canis familiaris
US-09-535-521-26

  Query Match             8.4%;  Score 62.8;  DB 4;  Length 369;
  Best Local Similarity   57.7%;  Pred. No. 3.7e-10;
  Matches  112;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||  || |||| |||   |||   | ||| ||   ||   | | | ||| ||| ||||
Db        205 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 146

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |     ||   |  ||    |||| ||| |  ||||||      || || ||||||||
Db        145 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 86

Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
              |  ||     |  || || || ||  || | | |      || ||||||||||||| |
Db         85 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 26

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db         25 CTGCGGCAGCTCGC 12


RESULT 10
US-09-535-521-7
; Sequence 7, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 384
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(384)
US-09-535-521-7

  Query Match             8.4%;  Score 62.8;  DB 4;  Length 384;
  Best Local Similarity   57.7%;  Pred. No. 3.8e-10;
  Matches  112;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||  || |||| |||   |||   | ||| ||   ||   | | | ||| ||| ||||
Db        156 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 215

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |     ||   |  ||    |||| ||| |  ||||||      || || ||||||||
Db        216 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 275

Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
              |  ||     |  || || || ||  || | | |      || ||||||||||||| |
Db        276 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 335

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db        336 CTGCGGCAGCTCGC 349


RESULT 11
US-09-535-521-9/c
; Sequence 9, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 384
;   TYPE: DNA
;   ORGANISM: Canis familiaris
US-09-535-521-9

  Query Match             8.4%;  Score 62.8;  DB 4;  Length 384;
  Best Local Similarity   57.7%;  Pred. No. 3.8e-10;
  Matches  112;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||  || |||| |||   |||   | ||| ||   ||   | | | ||| ||| ||||
Db        229 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 170

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |     ||   |  ||    |||| ||| |  ||||||      || || ||||||||
Db        169 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 110

Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
              |  ||     |  || || || ||  || | | |      || ||||||||||||| |
Db        109 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 50

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db         49 CTGCGGCAGCTCGC 36


RESULT 12
US-09-535-521-10
; Sequence 10, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
```

```
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 417
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(417)
US-09-535-521-10

  Query Match             8.4%;  Score 62.8;  DB 4;  Length 417;
  Best Local Similarity   57.7%;  Pred. No. 4e-10;
  Matches  112;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||  || |||| |||   |||   | ||| ||   ||   | | | ||| ||| ||||
Db        189 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 248

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |    ||   |  ||    |||| ||| |  ||||||      || || ||||||||
Db        249 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 308

Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
              |  ||    |  || || || ||  || | | |      || |||||||||||||| |
Db        309 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 368

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db        369 CTGCGGCAGCTCGC 382


RESULT 13
US-09-535-521-12/c
; Sequence 12, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 417
;   TYPE: DNA
;   ORGANISM: Canis familiaris
US-09-535-521-12

  Query Match             8.4%;  Score 62.8;  DB 4;  Length 417;
  Best Local Similarity   57.7%;  Pred. No. 4e-10;
  Matches  112;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||  || |||| |||   |||   | ||| ||   ||   | | | ||| ||| ||||
Db        229 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 170

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |    ||   |  ||    |||| ||| |  ||||||      || || ||||||||
Db        169 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 110

Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
              |  ||    |  || || || ||  || | | |      || |||||||||||||| |
Db        109 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 50

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db         49 CTGCGGCAGCTCGC 36


RESULT 14
US-09-535-521-13
; Sequence 13, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 423
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(423)
US-09-535-521-13

  Query Match             8.4%;  Score 62.8;  DB 4;  Length 423;
  Best Local Similarity   57.7%;  Pred. No. 4e-10;
  Matches  112;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||  || |||| |||   |||   | ||| ||   ||   | | | ||| ||| ||||
Db        195 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 254

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |    ||   |  ||    |||| ||| |  ||||||      || || ||||||||
Db        255 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 314

Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
              |  ||    |  || || || ||  || | | |      || |||||||||||||| |
Db        315 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 374

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db        375 CTGCGGCAGCTCGC 388


RESULT 15
US-09-535-521-15/c
; Sequence 15, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
```

; SEQ ID NO 15
; LENGTH: 423
; TYPE: DNA
; ORGANISM: Canis familiaris
US-09-535-521-15

Query Match 8.4%; Score 62.8; DB 4; Length 423;
Best Local Similarity 57.7%; Pred. No. 4e-10;
Matches 112; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy        525 GAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTT 584
              | ||  || |||| |||   |||   | ||| ||   ||   | | | ||| ||| ||||
Db        229 GTATGCCAACAAGAAGGGCACCTGGATTGGCCTCCGGGACCTGGACAGAGAGGGGGAGTT 170

Qy        585 TGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAA 644
              | |     ||   |  ||    |||| ||| |  ||||||      || || ||||||||
Db        169 TATCTGGATGGACGAGAACCCCCTGAACTATAGCAACTGGCGGCCCGGGGAGCCCAACAA 110

Qy        645 TGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCC 704
               |  ||    |  || || || ||  || | | |      || ||||||||||||| |
Db        109 CGGGGGCCAGGGCGAGGACTGCGTGATGATGCAGGGCTCGGGGCAGTGGAATGACGCCTT 50

Qy        705 CTGCTCCACCTCCC 718
              ||||  || ||| |
Db         49 CTGCGGCAGCTCGC 36
```

Search completed: June 21, 2004, 19:07:32
Job time : 82 secs

Blank

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 21, 2004, 18:23:44 ; Search time 409 Seconds
(without alignments)
8366.928 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 747
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	747	100.0	747	15	US-10-054-536-2	Sequence 2, Appli
2	747	100.0	3605	9	US-09-880-107-3705	Sequence 3705, Ap
3	747	100.0	3605	15	US-10-054-536-1	Sequence 1, Appli
4	747	100.0	3605	17	US-10-429-160-57	Sequence 57, Appl
5	684	91.6	684	13	US-09-971-475-2	Sequence 2, Appli
6	444.6	59.5	1409	15	US-10-076-816-54	Sequence 54, Appl
7	364.6	48.8	1037	9	US-09-917-800A-1710	Sequence 1710, Ap
8	287.6	38.5	1068	16	US-10-388-934-107	Sequence 107, App
9	222	29.7	419	9	US-09-960-352-11785	Sequence 11785, A
10	186	24.9	1802	13	US-10-070-415A-45	Sequence 45, Appl
11	186	24.9	1802	13	US-10-070-415A-49	Sequence 49, Appl
12	186	24.9	1802	13	US-10-070-415A-53	Sequence 53, Appl
13	185	24.8	1802	13	US-10-070-415A-41	Sequence 41, Appl
14	185	24.8	1802	13	US-10-070-415A-42	Sequence 42, Appl
15	185	24.8	1802	13	US-10-070-415A-43	Sequence 43, Appl
16	185	24.8	1802	13	US-10-070-415A-44	Sequence 44, Appl
17	184.4	24.7	1802	13	US-10-070-415A-46	Sequence 46, Appl
18	184.4	24.7	1802	13	US-10-070-415A-47	Sequence 47, Appl
19	184.4	24.7	1802	13	US-10-070-415A-48	Sequence 48, Appl
20	184.4	24.7	1802	13	US-10-070-415A-50	Sequence 50, Appl
21	184.4	24.7	1802	13	US-10-070-415A-51	Sequence 51, Appl
22	184.4	24.7	1802	13	US-10-070-415A-52	Sequence 52, Appl
23	184.4	24.7	1802	13	US-10-070-415A-54	Sequence 54, Appl
24	184.4	24.7	1802	13	US-10-070-415A-55	Sequence 55, Appl
25	184.4	24.7	1802	13	US-10-070-415A-56	Sequence 56, Appl
26	131.6	17.6	351	9	US-09-960-352-10110	Sequence 10110, A
c 27	112.4	15.0	5815	15	US-10-311-455-1208	Sequence 1208, Ap
28	108	14.5	542	16	US-10-388-934-108	Sequence 108, App
29	108	14.5	5815	15	US-10-311-455-1207	Sequence 1207, Ap
30	107.2	14.4	240	9	US-09-960-352-14616	Sequence 14616, A
31	100	13.4	140	13	US-09-727-030C-1	Sequence 1, Appli
32	95.6	12.8	221	9	US-09-960-352-2772	Sequence 2772, Ap
33	94.6	12.7	1313	13	US-10-116-802-47	Sequence 47, Appl
34	94.6	12.7	1410	9	US-09-954-456-2274	Sequence 2274, Ap
35	62.6	8.4	2248	9	US-09-925-302-30	Sequence 30, Appl
36	62.6	8.4	2248	13	US-09-925-302-30	Sequence 30, Appl
37	61.8	8.3	1161	12	US-10-336-603A-99	Sequence 99, Appl
38	61.4	8.2	752	12	US-10-336-603A-95	Sequence 95, Appl
39	61.4	8.2	935	10	US-09-971-429B-3	Sequence 3, Appli
40	60.2	8.1	4778	9	US-09-954-456-2031	Sequence 2031, Ap
41	58.6	7.8	681	12	US-10-336-603A-97	Sequence 97, Appl
42	57	7.6	4732	9	US-09-954-456-2079	Sequence 2079, Ap
43	57	7.6	4732	10	US-09-960-706-699	Sequence 699, App
44	57	7.6	4732	10	US-09-873-319-441	Sequence 441, App
45	52.4	7.0	544	15	US-10-002-631C-202	Sequence 202, App

ALIGNMENTS

RESULT 1
US-10-054-536-2
; Sequence 2, Application US/10054536
; Publication No. US20030162248A1
; GENERAL INFORMATION:
; APPLICANT: Wakamiya, No. US20030162248Alutaka
; TITLE OF INVENTION: RECOMBINANT HUMAN MANNAN-BINDING PROTEINS AND PROCESS
; TITLE OF INVENTION: FOR PRODUCING THE SAME
; FILE REFERENCE: 19036/36614
; CURRENT APPLICATION NUMBER: US/10/054,536
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03311
; PRIOR FILING DATE: 1998-07-23
; PRIOR APPLICATION NUMBER: JP 10-11864
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-054-536-2

Query Match 100.0%; Score 747; DB 15; Length 747;
Best Local Similarity 100.0%; Pred. No. 2.1e-236;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
```

RESULT 2
US-09-880-107-3705
; Sequence 3705, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3705
; LENGTH: 3605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X15422
US-09-880-107-3705

Query Match 100.0%; Score 747; DB 9; Length 3605;
Best Local Similarity 100.0%; Pred. No. 4.5e-236;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812
```

RESULT 3
US-10-054-536-1
; Sequence 1, Application US/10054536
; Publication No. US20030162248A1
; GENERAL INFORMATION:
; APPLICANT: Wakamiya, No. US20030162248Alutaka
; TITLE OF INVENTION: RECOMBINANT HUMAN MANNAN-BINDING PROTEINS AND PROCESS
; TITLE OF INVENTION: FOR PRODUCING THE SAME
; FILE REFERENCE: 19036/36614
; CURRENT APPLICATION NUMBER: US/10/054,536
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/JP98/03311
; PRIOR FILING DATE: 1998-07-23
; PRIOR APPLICATION NUMBER: JP 10-11864
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3605

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (66)..(809)
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: (126)..(809)
US-10-054-536-1

  Query Match             100.0%;  Score 747;  DB 15;  Length 3605;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-236;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812


RESULT 4
US-10-429-160-57
; Sequence 57, Application US/10429160
; Publication No. US20040023276A1
; GENERAL INFORMATION:
;  APPLICANT: Ward, Teresa R
;  APPLICANT: Mao, Mao
;  APPLICANT: Linsley, Peter S
;  APPLICANT: Lund, Lund
;  TITLE OF INVENTION: LXR Ligand Induced Genes and Proteins
;  FILE REFERENCE: RS0200
;  CURRENT APPLICATION NUMBER: US/10/429,160
;  CURRENT FILING DATE:  2003-05-02
;  PRIOR APPLICATION NUMBER: US 60/377,714
;  PRIOR FILING DATE: 2002-05-03
;  NUMBER OF SEQ ID NOS: 111
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
;   LENGTH: 3605
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-429-160-57

  Query Match             100.0%;  Score 747;  DB 17;  Length 3605;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-236;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812
```

```
RESULT 5
US-09-971-475-2
; Sequence 2, Application US/09971475
; Publication No. US20020086817A1
; GENERAL INFORMATION:
;  APPLICANT: Kawasaki, Toshisuke
;  TITLE OF INVENTION: Anticancer agent
;  FILE REFERENCE: ADT 308
;  CURRENT APPLICATION NUMBER: US/09/971,475
;  CURRENT FILING DATE:  2001-10-04
;  PRIOR APPLICATION NUMBER: US 09/468,705
;  PRIOR FILING DATE: 1999-12-21
;  PRIOR APPLICATION NUMBER: PCT/JP98/03697
;  PRIOR FILING DATE: 1998-08-19
;  PRIOR APPLICATION NUMBER: JP 239113/97
;  PRIOR FILING DATE: 1997-08-21
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 684
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-971-475-2

  Query Match             91.6%;  Score 684;  DB 13;  Length 684;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-215;
  Matches  684;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 60

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 120

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAGTTGGGGCCTCCAGGA 180

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 240

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 300

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 360

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 420

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 480

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 540

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 600

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 660

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              ||||||||||||||||||||||||
Db        661 CTGGCCGTCTGTGAGTTCCCTATC 684


RESULT 6
US-10-076-816-54
; Sequence 54, Application US/10076816
; Publication No. US20030056244A1
; GENERAL INFORMATION:
;  APPLICANT: Huang, Ning
;  APPLICANT:  Rodrguez, Raymond
;  APPLICANT:  Hagie, Frank E.
;  TITLE OF INVENTION: Feed Additive Compositions and Methods
;  FILE REFERENCE: 50665-8021.US00
;  CURRENT APPLICATION NUMBER: US/10/076,816
;  CURRENT FILING DATE:  2002-02-14
;  PRIOR APPLICATION NUMBER: US 60/269,188
;  PRIOR FILING DATE: 2001-02-14
;  PRIOR APPLICATION NUMBER: US 09/847,232
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/266,929
;  PRIOR FILING DATE: 2001-02-06
;  PRIOR APPLICATION NUMBER: US 60/201,182
;  PRIOR FILING DATE: 2000-05-02
;  NUMBER OF SEQ ID NOS: 60
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 54
;   LENGTH: 1409
;   TYPE: DNA
;   ORGANISM: Bos taurus
US-10-076-816-54

  Query Match             59.5%;  Score 444.6;  DB 15;  Length 1409;
  Best Local Similarity   76.4%;  Pred. No. 4.1e-136;
  Matches  575;  Conservative    0;  Mismatches  169;  Indels    9;  Gaps    2;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTT----- 55
              ||||| |||||| ||||||| ||| | |||||||||| |  |||| |||| ||||
Db         16 ATGTCGCTGTTTACATCACTTCCTTTCTTCTCCTGACTGCGGTGACAGCATCTTGTGCA 75

Qy         56 -ACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGT 114
               || ||||||| | || ||||||| |   ||  |||||||||| | ||||||||||||
Db         76 GACACAGAAACAGAGAACTGTGAGAACATCCGGAAGACCTGCC---CCGTGATTGCCTGT 132

Qy        115 AGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAA 174
              |  |||| |||||||| ||| ||||||||||||||||||||||  ||||||||||||
Db        133 GGTCCTCCGGGCATCAATGGCATCCCAGGCAAAGATGGGCGTGATGGTGCCAAGGGAGAA 192

Qy        175 AAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCT 234
              ||||| |||||||| ||||| |||||||||| |||||||||||||||||| ||||||||
Db        193 AAGGGAGAACCAGGTCAAGGACTCAGAGGCTCGCAGGGCCCCCCTGGAAAGATGGGGCCT 252

Qy        235 CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA 294
              | |||||  ||||||    |||||  |||||||||||  ||||||||||||||||||||
Db        253 CAAGGAACGCCAGGGATCCCTGGGATACCAGGACCAATAGGCCAAAAAGGAGACCCTGGA 312

Qy        295 AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
              |||  |  ||  ||   ||| | |||||||| | |||||||||  | ||||  ||| | |||
Db        313 GAAAATATGGGTGACTATATTCGCCTGGCTACTTCAGAAAGAGCAACTCTACAATCTGAA 372

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
              ||   |  |||||||| |||||  | ||||||||||||||||  |||||||||  |||
Db        373 TTGAACCAGATCAAAAACTGGCTAATCTTCTCTCTGGGCAAAAGAGTTGGGAAGAAGGCA 432

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              ||  | |||||||| |||  ||| | ||| |  |||||||  |  ||||||    |||||
Db        433 TTTTTTACCAATGGTAAAAAGATGCCTTTTAATGAAGTGAAGACTCTGTGTGCACAGTTC 492

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
              |||| |  |||||||||||| | |||||||| ||| ||  | |||  |  || ||  ||
Db        493 CAGGGCCGTGTGGCCACCCCTATGAATGCTGAAGAAAACAGGGCCCTCAAGGATTTAGTC 552
```

```
Qy        535 AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTG 594
              |  ||  || |||||||||||||||||| ||| || |||| |||||  | ||||||||||||
Db        553 ACTGAAGAGGCCTTCCTGGGCATCACAGATCAGGAGACTGAAGGCAAATTTGTGGATCTG 612

Qy        595 ACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCT 654
              ||||||||  |  ||||||||  |||||||||  ||  ||  ||  |||||  |||  | ||
Db        613 ACAGGAAAGGGGGTGACCTACCAAAACTGGAATGATGGCGAGCCTAACAACGCTTCTCCT 672

Qy        655 GATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACC 714
              |  ||  | |||||    || |||    | |||   |||||||| || | || ||| ||
Db        673 GGGGAGCACTGTGTGACACTTCTGTCGGACGGCACATGGAATGACATCGCTTGTTCCGCC 732

Qy        715 TCCCATCTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||  | ||  |||||||||| ||| ||  |||||
Db        733 TCCTTTTTGACCGTCTGTGAATTCTCTCTCTGA 765


RESULT 7
US-09-917-800A-1710
; Sequence 1710, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1710
;   LENGTH: 1037
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_022704
US-09-917-800A-1710

  Query Match             48.8%;  Score 364.6;  DB 9;  Length 1037;
  Best Local Similarity   70.7%;  Pred. No. 1.1e-109;
  Matches  483;  Conservative    0;  Mismatches  200;  Indels    0;  Gaps    0;

Qy         65 CTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAG 124
              | | |||||     || |      | |   ||    | |||||||||| || ||||| |
Db        233 CCGAGACCTTAACCGAAGGGGCTCAAAGTAGCTGCCCTGTGATTGCCTGCAGTTCTCCGG 292

Qy        125 GCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAAC 184
              |  |||||||||||||||||||||| |  || ||  |||||||||||||| ||||
Db        293 CCTGGAACGGCTTCCCAGGCAAAGATGGACACGACGGTGCCAAGGGAGAAAAGGGAGAAC 352

Qy        185 CAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATC 244
```

```
Db        353 CGGGTCAAGGCCTCAGAGGCTTGCAGGGCCCTCCTGGAAAAGTAGGACCTGCAGGGCCCC 412

Qy        245 CAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGG 304
              |||||  | ||||||||  |||| ||| ||| | ||||||||||| |||| | |||  |
Db        413 CAGGGAATCCTGGGTCAAAAGGAGCAACGGGACCAAAAGGAGACCGTGGAGAGAGTGTAG 472

Qy        305 ATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTA 364
              |  |||| || |   ||  | |||||   ||| |||| | | ||  ||  |  ||
Db        473 AATTTGATACTACCAACATTGATTTAGAAATTGCAGCCCTGCGATCGGAGCTGAGAGCTA 532

Qy        365 TCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCA 424
              | | ||||||| ||  | | ||| || |  || | ||||| || |||| |||| ||| ||
Db        533 TGAGAAAGTGGGTGCTCCTTTCTATGAGTGAAAATGTTGGAAAGAAGTACTTCATGAGCA 592

Qy        425 ATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTG 484
              || |  ||  ||| || ||  | | || ||||||  ||||   | |  ||||||  | |||
Db        593 GTGTTAGAAGGATGCCCCTTAACAGAGCGAAGGCTCTGTGCTCCGAACTCCAGGGCACTG 652

Qy        485 TGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAG 544
              |||||||  |||||||||||||||  || ||| | ||||| ||||||  |||  ||| |  |
Db        653 TGGCCACTCCCAGGAATGCTGAGGAAAATAGGGCCATCCAGAATGTGGCCAAAGATGTTG 712

Qy        545 CCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATA 604
              |||||  ||||||| ||  ||  ||| ||| |||      |||| ||| ||||||||||| |
Db        713 CCTTCTTGGGCATAACGGACCAGAGGACTGAAAACGTTTTTGAGGACCTGACAGGAAACA 772

Qy        605 GACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATT 664
              ||  ||  |||||| |||||||| |||||||| |||||||||  ||  ||||  ||| | |
Db        773 GAGTGCGCTACACTAACTGGAATGAGGGTGAGCCCAACAATGTGGGCTCTGGGGAAAACT 832

Qy        665 GTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGG 724
              ||||  ||||  ||| ||||||  |||||||||||| ||  |||||    |||  | |||
Db        833 GTGTGGTGCTCTTGACAAATGGGAAGTGGAATGACGTTCCTTGCTCTGATTCCTTTTTGG 892

Qy        725 CCGTCTGTGAGTTCCCTATCTGA 747
                ||  ||||| |||  ||  ||||
Db        893 TAGTTTGTGAATTCTCTGACTGA 915


RESULT 8
US-10-388-934-107
; Sequence 107, Application US/10388934
; Publication No. US20040005547A1
; GENERAL INFORMATION:
;  APPLICANT: Boess, Franziska
;  APPLICANT:  Suter-Dick, Laura
;  APPLICANT:  Wolf, Detlef
;  TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
;  FILE REFERENCE: 21199
;  CURRENT APPLICATION NUMBER: US/10/388,934
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 02005336.9
;  PRIOR FILING DATE: 2002-03-14
;  PRIOR APPLICATION NUMBER: 02015657.6
;  PRIOR FILING DATE: 2002-07-17
;  NUMBER OF SEQ ID NOS: 862
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 107
;   LENGTH: 1068
;   TYPE: DNA
;   ORGANISM: Rattus sp.
US-10-388-934-107

  Query Match             38.5%;  Score 287.6;  DB 16;  Length 1068;
  Best Local Similarity   68.1%;  Pred. No. 3.6e-84;
  Matches  417;  Conservative    0;  Mismatches  189;  Indels    6;  Gaps    1;

Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
              ||||  ||| ||| | |||||  |||| || || ||||| |||||||| ||||||||||
Db         64 GGCAGAGACGGGAGAGATGGGCCCAAAGGGGAGAAGGGAGAACCAGGTCAAGGGCTCAGG 123

Qy        202 GGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCA 261
```

```
Db        124 GGCTTGCAGGGCCCTCCAGGGAAACTGGGGCCTCCAGGAAGTGTAGGAGCCCCTGGAAGT 183

Qy        262 CCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTG 321

Db        184 CAAGGACCAAAAGGCCAAAAAGGGGATCGTGGAGACAGCAGAGCCATTGAGGTGAAGCTG 243

Qy        322 GC------TGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGG 375

Db        244 GCAAATATGGAGGCAGAGATAAACACCCTGAAGTCAAAACTGGAGCTAACCAACAAGTTG 303

Qy        376 CTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATA 435

Db        304 CATGCCTTCTCCATGGGTAAAAAGTCTGGGAAGAAGTTCTTTGTGACCAACCATGAAAGG 363

Qy        436 ATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCC 495

Db        364 ATGCCCTTTTCCAAAGTCAAGGCCCTGTGCTCAGAGCTCCGAGGCACTGTGGCTATCCCC 423

Qy        496 AGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGC 555

Db        424 AGGAATGCTGAGGAGAACAAGGCCATCCAAGAAGTGGCTAAAACCTCTGCCTTCCTAGGC 483

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615

Db        484 ATCACGGACGAGGTGACTGAAGGCCAATTCATGTATGTGACAGGGGGGAGGCTCACCTAC 543

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675

Db        544 AGCAACTGGAAAAAGGATGAGCCCAATGACCATGGCTCTGGGGAAGACTGTGTCACTATA 603

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735

Db        604 GTAGACAACGGTCTGTGGAATGACATCTCCTGCCAAGCTTCCCACACGGCTGTCTGCGAG 663

Qy        736 TTCCCTATCTGA 747

Db        664 TTCCCAGCCTGA 675


RESULT 9
US-09-960-352-11785
; Sequence 11785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11785
;   LENGTH: 419
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 50-LIB34-061-Q1-E1-E6
US-09-960-352-11785

  Query Match             29.7%;  Score 222;  DB 9;  Length 419;
  Best Local Similarity   78.3%;  Pred. No. 1.2e-62;
  Matches  307;  Conservative    0;  Mismatches   75;  Indels   10;  Gaps    3;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTT----- 55

Db         31 ATGTCGCTGTTTACATCACTTCCTTTTCTTCTCCTGACTGCGGTGACAGCATCTTGTGCA 90

Qy         56 -ACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGT 114

Db         91 GACACAGAAACAGAGAACTGTGAGAACATCCGGAAGACCTGCC---CCGTGATTGCCTGT 147

Qy        115 AGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAA 174

Db        148 GGTCCTCCGGGCATCAATGGCATCCCAGGCAAAGATGGGCGTGATGGTGCCAAGGGAGAA 207

Qy        175 AAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCT 234

Db        208 AAGGGAGAACCAGGTCAAGGACTCAGAGGCTCGCAGGGCCCCCCTGGAAAGATGGGGCCT 267

Qy        235 CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA 294

Db        268 CAAGGAACGCCAGGGATCCCTGGGATACCAGGACCAATAGGCCAAAAAGGAGACCCTGGA 327

Qy        295 AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAA-GAAAAGCTCTGCAAACAGA 353

Db        328 GAAAATATGGGTGACTATATTCGCCTGGCTACCTTAAAAAGGAGCAATTCTACAATCTGA 387

Qy        354 AATGGCACGTATCAAAAAGTGGCTGACCTTCT 385

Db        388 ATTGAACCAGATCAAAAACTGGCTAATCTTCT 419


RESULT 10
US-10-070-415A-45
; Sequence 45, Application US/10070415A
; Publication No. US20040043379A1
; GENERAL INFORMATION:
;  APPLICANT: HASHIMOTO, Koji
;  APPLICANT:  ASHIMOTO, Michie
;  APPLICANT:  MISHIRO, Shunji
;  APPLICANT:  OOTA, Yasuhiko
;  TITLE OF INVENTION: DETECTION OF NUCLEIC ACID ASSOCIATED WITH DISEASE
;  FILE REFERENCE: 220633US2SRDPCT
;  CURRENT APPLICATION NUMBER: US/10/070,415A
;  CURRENT FILING DATE:  2003-07-23
;  PRIOR APPLICATION NUMBER: PCT/JP02/02030
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: JP 2001-090053
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: JP 2001-284112
;  PRIOR FILING DATE: 2001-09-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
;   LENGTH: 1802
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (96)..(96)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (649)..(649)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (868)..(868)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (884)..(884)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (425)..(425)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
US-10-070-415A-45

  Query Match             24.9%;  Score 186;  DB 13;  Length 1802;
  Best Local Similarity   98.9%;  Pred. No. 2.1e-50;
  Matches  186;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 774

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 834

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db        835 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGNGTGATGGCACCAAGGNAGAAAAGGGG 894

Qy        181 GAACCAGG 188
              ||||||||
Db        895 GAACCAGG 902


RESULT 11
US-10-070-415A-49
; Sequence 49, Application US/10070415A
; Publication No. US20040043379A1
; GENERAL INFORMATION:
;  APPLICANT: HASHIMOTO, Koji
;  APPLICANT:  ASHIMOTO, Michie
;  APPLICANT:  MISHIRO, Shunji
;  APPLICANT:  OOTA, Yasuhiko
;  TITLE OF INVENTION: DETECTION OF NUCLEIC ACID ASSOCIATED WITH DISEASE
;  FILE REFERENCE: 220633US2SRDPCT
;  CURRENT APPLICATION NUMBER: US/10/070,415A
;  CURRENT FILING DATE:  2003-07-23
;  PRIOR APPLICATION NUMBER: PCT/JP02/02030
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: JP 2001-090053
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: JP 2001-284112
;  PRIOR FILING DATE: 2001-09-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
;   LENGTH: 1802
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (96)..(96)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (425)..(425)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (649)..(649)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (884)..(884)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (875)..(875)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
US-10-070-415A-49

  Query Match           24.9%;  Score 186;  DB 13;  Length 1802;
  Best Local Similarity   98.9%;  Pred. No. 2.1e-50;
  Matches  186;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 774

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 834

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db        835 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGNGTGATGGCACCAAGGNAGAAAAGGGG 894

Qy        181 GAACCAGG 188
              ||||||||
Db        895 GAACCAGG 902


RESULT 12
US-10-070-415A-53
; Sequence 53, Application US/10070415A
; Publication No. US20040043379A1
; GENERAL INFORMATION:
;  APPLICANT: HASHIMOTO, Koji
;  APPLICANT:  ASHIMOTO, Michie
;  APPLICANT:  MISHIRO, Shunji
;  APPLICANT:  OOTA, Yasuhiko
;  TITLE OF INVENTION: DETECTION OF NUCLEIC ACID ASSOCIATED WITH DISEASE
;  FILE REFERENCE: 220633US2SRDPCT
;  CURRENT APPLICATION NUMBER: US/10/070,415A
;  CURRENT FILING DATE:  2003-07-23
;  PRIOR APPLICATION NUMBER: PCT/JP02/02030
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: JP 2001-090053
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: JP 2001-284112
;  PRIOR FILING DATE: 2001-09-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
;   LENGTH: 1802
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (96)..(96)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (425)..(425)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (649)..(649)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (868)..(868)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (875)..(875)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
US-10-070-415A-53

  Query Match           24.9%;  Score 186;  DB 13;  Length 1802;
  Best Local Similarity   98.9%;  Pred. No. 2.1e-50;
  Matches  186;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 774

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 834

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||
```

```
Db        835 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGNGTGATGNCACCAAGGGAGAAAAGGGG 894

Qy        181 GAACCAGG 188
              ||||||||
Db        895 GAACCAGG 902


RESULT 13
US-10-070-415A-41
; Sequence 41, Application US/10070415A
; Publication No. US20040043379A1
; GENERAL INFORMATION:
;  APPLICANT: HASHIMOTO, Koji
;  APPLICANT:  ASHIMOTO, Michie
;  APPLICANT:  MISHIRO, Shunji
;  APPLICANT:  OOTA, Yasuhiko
;  TITLE OF INVENTION: DETECTION OF NUCLEIC ACID ASSOCIATED WITH DISEASE
;  FILE REFERENCE: 220633US2SRDPCT
;  CURRENT APPLICATION NUMBER: US/10/070,415A
;  CURRENT FILING DATE:  2003-07-23
;  PRIOR APPLICATION NUMBER: PCT/JP02/02030
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: JP 2001-090053
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: JP 2001-284112
;  PRIOR FILING DATE: 2001-09-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
;   LENGTH: 1802
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (96)..(96)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (649)..(649)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (868)..(868)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (875)..(875)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (884)..(884)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
US-10-070-415A-41

  Query Match             24.8%;  Score 185;  DB 13;  Length 1802;
  Best Local Similarity   98.4%;  Pred. No. 4.4e-50;
  Matches  185;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 774

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 834

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||| |||||| |||||||| ||||||||||
Db        835 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGNGTGATGNCACCAAGGNAGAAAAGGGG 894

Qy        181 GAACCAGG 188
              ||||||||
Db        895 GAACCAGG 902


RESULT 14
US-10-070-415A-42
; Sequence 42, Application US/10070415A
; Publication No. US20040043379A1
; GENERAL INFORMATION:
;  APPLICANT: HASHIMOTO, Koji
;  APPLICANT:  ASHIMOTO, Michie
;  APPLICANT:  MISHIRO, Shunji
;  APPLICANT:  OOTA, Yasuhiko
;  TITLE OF INVENTION: DETECTION OF NUCLEIC ACID ASSOCIATED WITH DISEASE
;  FILE REFERENCE: 220633US2SRDPCT
;  CURRENT APPLICATION NUMBER: US/10/070,415A
;  CURRENT FILING DATE:  2003-07-23
;  PRIOR APPLICATION NUMBER: PCT/JP02/02030
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: JP 2001-090053
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: JP 2001-284112
;  PRIOR FILING DATE: 2001-09-18
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
;   LENGTH: 1802
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (96)..(96)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (649)..(649)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (868)..(868)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (875)..(875)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (884)..(884)
;   OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
US-10-070-415A-42

  Query Match             24.8%;  Score 185;  DB 13;  Length 1802;
  Best Local Similarity   98.4%;  Pred. No. 4.4e-50;
  Matches  185;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 774

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 834

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||| |||||| |||||||| ||||||||||
Db        835 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGNGTGATGNCACCAAGGNAGAAAAGGGG 894

Qy        181 GAACCAGG 188
              ||||||||
Db        895 GAACCAGG 902


RESULT 15
US-10-070-415A-43
; Sequence 43, Application US/10070415A
```

; Publication No. US20040043379A1
; GENERAL INFORMATION:
; APPLICANT: HASHIMOTO, Koji
; APPLICANT: ASHIMOTO, Michie
; APPLICANT: MISHIRO, Shunji
; APPLICANT: OOTA, Yasuhiko
; TITLE OF INVENTION: DETECTION OF NUCLEIC ACID ASSOCIATED WITH DISEASE
; FILE REFERENCE: 220633US2SRDPCT
; CURRENT APPLICATION NUMBER: US/10/070,415A
; CURRENT FILING DATE: 2003-07-23
; PRIOR APPLICATION NUMBER: PCT/JP02/02030
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: JP 2001-090053
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: JP 2001-284112
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 1802
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (96)..(96)
; OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (649)..(649)
; OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (868)..(868)
; OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (875)..(875)
; OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (884)..(884)
; OTHER INFORMATION: n is a nucleotide selected from a, g, c, or t
US-10-070-415A-43

Query Match 24.8%; Score 185; DB 13; Length 1802;
Best Local Similarity 98.4%; Pred. No. 4.4e-50;
Matches 185; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 774

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 834

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||| |||||| |||||||| ||||||||||
Db        835 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGNGTGATGNCACCAAGGNAGAAAAGGGG 894

Qy        181 GAACCAGG 188
              ||||||||
Db        895 GAACCAGG 902
```

Search completed: June 21, 2004, 20:12:09
Job time : 412 secs

Blank

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 18, 2004, 20:53:21 ; Search time 15 Seconds
(without alignments)
5186.187 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 1365
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10054536/runat_18062004_173455_13573/app_query.fasta_1.903
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054536_@CGN_1_1_16_@runat_18062004_173455_13573 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1321	96.8	248	1	MABC_HUMAN	P11226 homo sapien
2	847.5	62.1	249	1	MABC_BOVIN	O02659 bos taurus
3	790	57.9	244	1	MABC_MOUSE	P41317 mus musculu
4	785	57.5	244	1	MABC_RAT	P08661 rattus norv
5	711.5	52.1	239	1	MABA_MOUSE	P39039 mus musculu
6	705.5	51.7	238	1	MABA_RAT	P19999 rattus norv
7	454	33.3	375	1	PSPD_HUMAN	P35247 homo sapien
8	449.5	32.9	371	1	CL46_BOVIN	Q8mhz9 bos taurus
9	441.5	32.3	371	1	CONG_BOVIN	P23805 bos taurus
10	441	32.3	374	1	PSPD_RAT	P35248 rattus norv
11	439	32.2	374	1	PSPD_MOUSE	P50404 mus musculu
12	427	31.3	321	1	CL43_BOVIN	P42916 bos taurus
13	417.5	30.6	247	1	PSPA_RABIT	P12842 oryctolagus
14	415.5	30.4	369	1	PSPD_BOVIN	P35246 bos taurus
15	405.5	29.7	247	1	PSPA_CAVPO	P50403 cavia porce
16	404	29.6	248	1	PSPA_RAT	P08427 rattus norv
17	393.5	28.8	248	1	PSPA_HUMAN	P07714 homo sapien
18	393	28.8	248	1	PSPA_MOUSE	P35242 mus musculu
19	393	28.8	249	1	PSPA_PIG	P49874 sus scrofa
20	374.5	27.4	248	1	PSPA_CANFA	P06908 canis famil
21	204	14.9	197	1	CLE1_HUMAN	O75596 homo sapien
22	201.5	14.8	331	1	FCE2_MOUSE	P20693 mus musculu
23	199.5	14.6	207	1	LECH_CHICK	P02707 gallus gall
24	193.5	14.2	253	1	C1QB_MOUSE	P14106 mus musculu
25	192.5	14.1	253	1	C1QB_RAT	P31721 rattus norv
26	192.5	14.1	335	1	FCN1_RAT	Q9wts8 rattus norv
27	192	14.1	244	1	APM1_HUMAN	Q15848 homo sapien
28	189.5	13.9	247	1	APM1_MOUSE	Q60994 mus musculu
29	189	13.8	245	1	C1QA_MOUSE	P98086 mus musculu
30	188	13.8	1670	1	CA34_HUMAN	Q01955 homo sapien
31	187	13.7	334	1	FCN1_MOUSE	O70165 mus musculu
32	187	13.7	1763	1	CA24_ASCSU	P27393 ascaris suu
33	187	13.7	1888	1	CA1E_CHICK	P32018 gallus gall
34	186	13.6	1690	1	CA44_HUMAN	P53420 homo sapien
35	185.5	13.6	202	1	TETN_HUMAN	P05452 homo sapien
36	185	13.6	912	1	PGCB_BOVIN	Q28062 bos taurus
37	184	13.5	251	1	C1QB_HUMAN	P02746 homo sapien
38	183.5	13.4	245	1	C1QA_HUMAN	P02745 homo sapien
39	183	13.4	675	1	CA39_CHICK	P32017 gallus gall
40	183	13.4	688	1	CA29_MOUSE	Q07643 mus musculu
41	182.5	13.4	245	1	C1QC_HUMAN	P02747 homo sapien
42	182.5	13.4	246	1	C1QC_MOUSE	Q02105 mus musculu
43	181.5	13.3	1329	1	KF10_HUMAN	Q9p218 homo sapien
44	181.5	13.3	1459	1	CA12_HUMAN	P28481 mus musculu
45	181	13.3	689	1	CA29_HUMAN	Q14055 homo sapien

ALIGNMENTS

RESULT 1
MABC_HUMAN
ID MABC_HUMAN STANDARD; PRT; 248 AA.
AC P11226; Q86SI4; Q96KE4; Q96TF7; Q96TF8; Q96TF9;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mannose-binding protein C precursor (MBP-C) (MBP1) (Mannan-binding
DE protein) (Mannose-binding lectin).
GN MBL2 OR MBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90010778; PubMed=2477486;
RA Sastry K., Herman G.A., Day L.E., Deignan E., Bruns G., Morton C.C.,
RA Ezekowitz R.A.B.;
RT "The human mannose-binding protein gene. Exon structure reveals its
RT evolutionary relationship to a human pulmonary surfactant gene and
RT localization to chromosome 10.";
RL J. Exp. Med. 170:1175-1189(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88171281; PubMed=2450948;
RA Ezekowitz R.A.B., Day L.E., Herman G.A.;
RT "A human mannose-binding protein is an acute-phase reactant that
RT shares sequence homology with other vertebrate lectins.";
RL J. Exp. Med. 167:1034-1046(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90073571; PubMed=2590164;
RA Taylor M.E., Brickell P.M., Craig R.K., Summerfield J.A.;
RT "Structure and evolutionary origin of the gene encoding a human serum
RT mannose-binding protein.";
RL Biochem. J. 262:763-771(1989).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANTS CYS-52; ASP-54 AND GLU-57.
RX MEDLINE=98414317; PubMed=9743385;
RA Madsen H.O., Satz M.L., Hogh B., Svejgaard A., Garred P.;

```
RT   "Different molecular events result in low protein levels of mannan-
RT   binding lectin in populations from South-East Africa and South
RT   America.";
RL   J. Immunol. 161:3169-3175(1998).
RN   [5]
RP   SEQUENCE FROM N.A., AND VARIANT ALA-24.
RC   TISSUE=Liver;
RA   Chen Z., Zhu X., Xie P.;
RT   "Cloning and sequencing of mannan-binding lectin cDNA of Chinese.";
RL   Mian Yi Xue Za Zhi 15:83-86(1999).
RN   [6]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Liver, and Plasma;
RX   MEDLINE=95073978; PubMed=7982896;
RA   Kurata H., Sannoh T., Kozutsumi Y., Yokota Y., Kawasaki T.;
RT   "Structure and function of mannan-binding proteins isolated from
RT   human liver and serum.";
RL   J. Biochem. 115:1148-1154(1994).
RN   [7]
RP   SEQUENCE OF 1-59 FROM N.A., AND VARIANT ASP-54.
RX   MEDLINE=22167090; PubMed=12175909;
RA   Jueliger S., Kremsner P.G., Alpers M.P., Reeder J.C., Kun J.F.J.;
RT   "Restricted polymorphisms of the mannose-binding lectin gene in a
RT   population of Papua New Guinea.";
RL   Mutat. Res. 505:87-91(2002).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 108-248.
RX   MEDLINE=95360730; PubMed=7634089;
RA   Sheriff S., Chang C.Y., Ezekowitz R.A.;
RT   "Human mannose-binding protein carbohydrate recognition domain
RT   trimerizes through a triple alpha-helical coiled-coil.";
RL   Nat. Struct. Biol. 1:789-794(1994).
RN   [9]
RP   VARIANT ASP-54.
RX   MEDLINE=91269930; PubMed=1675710;
RA   Sumiya M., Super M., Tabona P., Levinsky R.J., Arai T., Turner M.W.,
RA   Summerfield J.A.;
RT   "Molecular basis of opsonic defect in immunodeficient children.";
RL   Lancet 337:1569-1570(1991).
RN   [10]
RP   VARIANTS ASP-54 AND GLU-57.
RX   MEDLINE=93258313; PubMed=1304173;
RA   Lipscombe R.J., Sumiya M., Hill A.V.S., Lau Y.L., Levinsky R.J.,
RA   Summerfield J.A., Turner M.W.;
RT   "High frequencies in African and non-African populations of
RT   independent mutations in the mannose binding protein gene.";
RL   Hum. Mol. Genet. 1:709-715(1992).
RN   [11]
RP   ERRATUM.
RA   Lipscombe R.J., Sumiya M., Hill A.V.S., Lau Y.L., Levinsky R.J.,
RA   Summerfield J.A., Turner M.W.;
RL   Hum. Mol. Genet. 2:342-342(1993).
RN   [12]
RP   VARIANT ASP-54.
RX   MEDLINE=93265124; PubMed=1303250;
RA   Super M., Gillies S.D., Foley S., Sastry K., Schweinle J.E.,
RA   Silverman V.J., Ezekowitz R.A.;
RT   "Distinct and overlapping functions of allelic forms of human mannose
RT   binding protein.";
RL   Nat. Genet. 2:50-55(1992).
RN   [13]
RP   VARIANTS CYS-52; ASP-54 AND GLU-57.
RX   MEDLINE=99374928; PubMed=10447262;
RA   Gabolde M., Muralitharan S., Besmond C.;
RT   "Genotyping of the three major allelic variants of the human
RT   mannose-binding lectin gene by denaturing gradient gel
RT   electrophoresis.";
RL   Hum. Mutat. 14:80-83(1999).
CC   -!- FUNCTION: Binds mannose and N-acetylglucosamine in a calcium-
CC       dependent manner. Is capable of host defense against pathogens,
CC       by activating the classical complement pathway independently of
CC       the antibody.
CC   -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC   -!- DISEASE: There is an association between low levels of MBL2 and a
CC       defect of opsonization which results in susceptibility to frequent
CC       and chronic infections.
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- CAUTION: There are extensive differences between the revised
CC       sequence in Ref.1 and that published in Ref.2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15422; CAA33462.1; -.
DR   EMBL; X15954; CAA34079.1; -.
DR   EMBL; X15955; CAA34079.1; JOINED.
DR   EMBL; X15956; CAA34079.1; JOINED.
DR   EMBL; X15957; CAA34079.1; JOINED.
DR   EMBL; AF080510; AAC31937.1; -.
DR   EMBL; AF080508; AAC31937.1; JOINED.
DR   EMBL; AF080509; AAC31937.1; JOINED.
DR   EMBL; Y16576; CAB56044.1; -.
DR   EMBL; Y16577; CAB56120.1; -.
DR   EMBL; Y16578; CAB56045.1; -.
DR   EMBL; Y16579; CAB56121.1; -.
DR   EMBL; Y16580; CAB56122.1; -.
DR   EMBL; Y16581; CAB56123.1; -.
DR   EMBL; Y16582; CAB56124.1; -.
DR   EMBL; AF360991; AAK52907.1; -.
DR   EMBL; AF482699; AAN39274.1; -.
DR   EMBL; AF482700; AAN39275.1; -.
DR   PIR; JL0115; LNHUMC.
DR   PDB; 1HUP; 15-OCT-95.
DR   Genew; HGNC:6922; MBL2.
DR   MIM; 154545; -.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0003793; F:defense/immunity protein activity; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Complement pathway; Membrane; Mannose-binding; Calcium; Repeat;
KW   Signal; Collagen; Lectin; Glycoprotein; Hydroxylation; Polymorphism;
KW   3D-structure.
FT   SIGNAL       1    20
FT   CHAIN       21   248       MANNOSE-BINDING PROTEIN C.
FT   DOMAIN      21    41       CYS-RICH.
FT   DOMAIN      42    99       COLLAGEN-LIKE.
FT   DOMAIN     134   245       C-TYPE LECTIN.
FT   MOD_RES     47    47       HYDROXYLATION.
FT   MOD_RES     73    73       HYDROXYLATION.
FT   MOD_RES     79    79       HYDROXYLATION.
FT   MOD_RES     82    82       HYDROXYLATION.
FT   MOD_RES     88    88       HYDROXYLATION.
FT   DISULFID   155   244
FT   DISULFID   222   236
FT   VARIANT     24    24       T -> A (in Chinese).
FT                              /FTId=VAR_013294.
FT   VARIANT     52    52       R -> C (in 0.05% of European and African
FT                              populations; dbSNP:5030737).
FT                              /FTId=VAR_008543.
FT   VARIANT     54    54       G -> D (in Caucasian and Chinese
FT                              populations; dbSNP:1800450).
FT                              /FTId=VAR_004182.
FT   VARIANT     57    57       G -> E (in West African population;
FT                              dbSNP:1800451).
```

```
FT                                    /FTId=VAR_004183.
FT   TURN        109    109
FT   HELIX       110    129
FT   TURN        130    130
FT   STRAND      132    134
FT   TURN        135    136
FT   STRAND      137    147
FT   HELIX       148    157
FT   TURN        158    159
FT   STRAND      161    162
FT   HELIX       168    177
FT   STRAND      182    187
FT   TURN        192    193
FT   STRAND      196    197

Alignment Scores:
Pred. No.:              2.02e-102      Length:         248
Score:                  1321.00        Matches:        248
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            96.78%         Indels:         0
DB:                     1              Gaps:           0

US-10-054-536-2 (1-747) x MABC_HUMAN (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248

RESULT 2
MABC_BOVIN
ID   MABC_BOVIN     STANDARD;      PRT;   249 AA.
AC   O02659;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein).
GN   MBL.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97228413; PubMed=9074491;
RA   Kawai T., Suzuki Y., Eda S., Ohtani K., Kase T., Fujinaga Y.,
RA   Sakamoto T., Kurimura T., Wakamiya N.;
RT   "Cloning and characterization of a cDNA encoding bovine mannan-binding
RT   protein.";
RL   Gene 186:161-165(1997).
CC   -!- FUNCTION: Binds mannose and N-acetylglucosamine in a calcium-
CC       dependent manner. Is capable of host defense against pathogens, by
CC       activating the classical complement pathway independently of the
CC       antibody (By similarity).
CC   -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D73408; BAA18935.1; -.
DR   HSSP; P11226; 1HUP.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Complement pathway; Membrane; Mannose-binding; Calcium; Repeat;
KW   Signal; Collagen; Lectin; Glycoprotein; Hydroxylation.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    249       MANNOSE-BINDING PROTEIN C.
FT   DOMAIN       43    101       COLLAGEN-LIKE.
FT   DOMAIN      135    246       C-TYPE LECTIN.
FT   MOD_RES      48     48       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      63     63       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      74     74       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      83     83       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      86     86       HYDROXYLATION (POTENTIAL).
FT   DISULFID     34     34       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     39     39       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    156    245       BY SIMILARITY.
FT   DISULFID    223    237       BY SIMILARITY.
SQ   SEQUENCE   249 AA;  26471 MW;  29FC9F5927A66DD5 CRC64;

Alignment Scores:
Pred. No.:              5.88e-63       Length:         249
```

Score:	847.50	Matches:	161
Percent Similarity:	75.20%	Conservative:	27
Best Local Similarity:	64.40%	Mismatches:	59
Query Match:	62.09%	Indels:	3
DB:	1	Gaps:	2

US-10-054-536-2 (1-747) x MABC_BOVIN (1-249)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||   |||||||||   |||||||||:::   |||   ||||||   :::
Db          1 MetSerLeuPheThrSerLeuProPheLeuLeuLeuThrAlaValThrAlaSerCysAla 20

Qy         61 GAAACTGTGACC------TGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGT 114
              :::|||   |||      ||||||:::   :::||||||||||||   ||||||||||||
Db         21 AspThrGluThrGluAsnCysGluAsnIleArgLysThrCysPro---ValIleAlaCys 39

Qy        115 AGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAA 174
              ||||||||||||||   |||||||||||||||||||||||||   |||||||||
Db         40 GlyProProGlyIleAsnGlyIleProGlyLysAspGlyArgAspGlyAlaLysGlyGlu 59

Qy        175 AAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCT 234
              |||||||||||||||||||||||||||   |||||||||||||||||||:::||||||
Db         60 LysGlyGluProGlyGlnGlyLeuArgGlySerGlnGlyProProGlyLysMetGlyPro 79

Qy        235 CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA 294
                 |||   ||||||      |||   |||||||||   ||||||||||||||||||||
Db         80 GlnGlyThrProGlyIleProGlyIleProGlyProIleGlyGlnLysGlyAspProGly 99

Qy        295 AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
              ::::::                  ||||||   |||||||||      ||||||:::|||
Db        100 GluAsnMetGlyAspTyrIleArgLeuAlaThrSerGluArgAlaThrLeuGlnSerGlu 119

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
              :::   :::||||||   ||||||   ||||||||||||||:::||||||   |||
Db        120 LeuAsnGlnIleLysAsnTrpLeuIlePheSerLeuGlyLysArgValGlyLysLysAla 139

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              |||   |||||||||:::   |||   |||   :::||||||   ||||||   :::|||
Db        140 PhePheThrAsnGlyLysLysMetProPheAsnGluValLysThrLeuCysAlaGlnPhe 159

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
              |||      ||||||||||||   ||||||   ||||||   |||:::::::::|||:::
Db        160 GlnGlyArgValAlaThrProMetAsnAlaGluGluAsnArgAlaLeuLysAspLeuVal 179

Qy        535 AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTG 594
              |||||||||||||||||||||||||:::::|||||||||:::||||||||||||
Db        180 ThrGluGluAlaPheLeuGlyIleThrAspGlnGluThrGluGlyLysPheValAspLeu 199

Qy        595 ACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCT 654
              ||||||      :::||||||   |||||||||:::|||||||||||||||||||
Db        200 ThrGlyLysGlyValThrTyrGlnAsnTrpAsnAspGlyGluProAsnAsnAlaSerPro 219

Qy        655 GATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACC 714
              |||   ||||||   ||||||   :::|||   |||||||||:::   ||||||
Db        220 GlyGluHisCysValThrLeuLeuSerAspGlyThrTrpAsnAspIleAlaCysSerAla 239

Qy        715 TCCCATCTGGCCGTCTGTGAGTTCCCTATC 744
              |||   |||   ||||||||||||   :::
Db        240 SerPheLeuThrValCysGluPheSerLeu 249
```

RESULT 3
MABC_MOUSE
ID MABC_MOUSE STANDARD; PRT; 244 AA.
AC P41317;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)
DE (RA-reactive factor P28A subunit) (RARF/P28A).
GN MBL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBA/J; TISSUE=Liver;
RX MEDLINE=91302823; PubMed=1712818;
RA Sastry K., Zahedi K., Lelias J.M., Whitehead A.S., Ezekowitz R.A.;
RT "Molecular characterization of the mouse mannose-binding proteins.
RT The mannose-binding protein A but not C is an acute phase reactant.";
RL J. Immunol. 147:692-697(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=95284466; PubMed=7766991;
RA Sastry R., Wang J.S., Brown D.C., Ezekowitz R.A., Tauber A.I.,
RA Sastry K.N.;
RT "Characterization of murine mannose-binding protein genes Mbl1 and
RT Mbl2 reveals features common to other collectin genes.";
RL Mamm. Genome 6:103-110(1995).
RN [3]
RP SEQUENCE FROM N.A.
RA Kuge S., Ihara S., Watanabe E., Watanabe M., Takishima K., Suga T.,
RA Mamaiya G., Kawakami M.;
RL Submitted (JUL-1992) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Binds mannose and N-acetylglucosamine in a calcium-
CC dependent manner. Is capable of host defense against pathogens, by
CC activating the classical complement pathway independently of the
CC antibody.
CC -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST
CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL
CC SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE
CC PROTEIN.
CC -!- SIMILARITY: Contains 1 collagenous domain.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC --

CC --
DR EMBL; S42294; AAB19343.1; -.
DR EMBL; U09016; AAA82010.1; -.
DR EMBL; U09013; AAA82010.1; JOINED.
DR EMBL; U09014; AAA82010.1; JOINED.

```
DR   EMBL; U09015; AAA82010.1; JOINED.
DR   EMBL; D11440; BAA02005.1; -.
DR   EMBL; BC010760; AAH10760.1; -.
DR   PIR; I48651; LNMSMC.
DR   HSSP; P08661; 1RDO.
DR   MGD; MGI:96924; Mbl2.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Complement pathway; Membrane; Mannose-binding; Calcium; Repeat;
KW   Signal; Collagen; Lectin; Glycoprotein; Hydroxylation.
FT   SIGNAL        1     18       BY SIMILARITY.
FT   CHAIN        19    244       MANNOSE-BINDING PROTEIN C.
FT   DOMAIN       38     96       COLLAGEN-LIKE.
FT   DOMAIN      129    241       C-TYPE LECTIN.
FT   MOD_RES      43     43       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      58     58       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      69     69       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      78     78       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      81     81       HYDROXYLATION (POTENTIAL).
FT   DISULFID     29     29       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     34     34       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    151    240       BY SIMILARITY.
FT   DISULFID    218    232       BY SIMILARITY.
FT   CONFLICT      3      3       I -> L (IN REF. 1).
FT   CONFLICT     15     15       V -> A (IN REF. 1).
SQ   SEQUENCE   244 AA;  25957 MW;  49AE84E2290DEB0A CRC64;

Alignment Scores:
Pred. No.:                3.63e-58      Length:          244
Score:                    790.00        Matches:         150
Percent Similarity:       75.90%        Conservative:    39
Best Local Similarity:    60.24%        Mismatches:      50
Query Match:              57.88%        Indels:          10
DB:                       1             Gaps:            5

US-10-054-536-2 (1-747) x MABC_MOUSE (1-244)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||:::|||   |||        |||||||||   :::|||         |||:::
Db          1 MetSerIlePheThrSer------PheLeuLeuLeuCysValValThrValValTyrAla 18

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||:::|||   |||       |||   :::||||||   |||:::   |||||||||
Db         19 GluThrLeuThr---GluGlyValGlnAsnSerCysPro---ValValThrCysSerSer 36

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||:::|||||||||||||||||||||||||||||||   |||||||||||||||
Db         37 ProGlyLeuAsnGlyPheProGlyLysAspGlyArgAspGlyAlaLysGlyGluLysGly 56

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTG---------GGG 231
              |||||||||||||||||||||||||||||||||||||||||:::         |||
Db         57 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysValGlyProThrGly 76

Qy        232 CCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCT 291
              ||||||||||||||||      |||:::   ||||||||||   :::||||||
Db         77 ProProGlyAsnProGlyLeuLysGlyAlaValGlyProLysGlyAspArgGlyAsp--- 95

Qy        292 GGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACA 351
                      :::   |||:::|||         ||||||      ||||||::::::
Db         96 ------ArgAlaGluPheAspThrSerGluIleAspSerGluIleAlaAlaLeuArgSer 113

Qy        352 GAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTCTGGGCAAACAAGTTGGGAACAAG 411
              |||:::      ::::::   |||:::   |||||||||   ::::::||||||   |||
Db        114 GluLeuArgAlaLeuArgAsnTrpValLeuPheSerLeuSerGluLysValGlyLysLys 133

Qy        412 TTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAG 471
              :::|||:::::::::   :::   |||:::   ::::::||||||||||||||   :::
Db        134 TyrPheValSerSerValLysLysMetSerLeuAspArgValLysAlaLeuCysSerGlu 153

Qy        472 TTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTC 531
              ||||||   ||||||||||||||||||||||   ||||||   ||||||||   :::
Db        154 PheGlnGlySerValAlaThrProArgAsnAlaGluGluAsnSerAlaIleGlnLysVal 173

Qy        532 ATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGAT 591
                 |||:::   |||:::|||||||||||||||   :::   ||||||   |||   |||
Db        174 AlaLysAspIleAlaTyrLeuGlyIleThrAspValArgValGluGlySerPheGluAsp 193

Qy        592 CTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGT 651
              ||||||||||||||:::   ||||||||||||||:::||||||||||||||   |||
Db        194 LeuThrGlyAsnArgValArgTyrThrAsnTrpAsnAspGlyGluProAsnAsnThrGly 213

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
                    ||||||||||||:::::|||   ||||||:::|||||||||||||||||||||
Db        214 AspGlyGluAspCysValValIleLeuGlyAsnGlyLysTrpAsnAspValProCysSer 233

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
                 |||   ||||||:::|||||||||
Db        234 AspSerPheLeuAlaIleCysGluPhe 242

RESULT 4
MABC_RAT
ID   MABC_RAT       STANDARD;      PRT;   244 AA.
AC   P08661;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)
DE   (RA-reactive factor P28A subunit) (RARF/P28A).
GN   MBL2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92299655; PubMed=1607365;
RA   Wada M., Itoh N., Ohta M., Kawasaki T.;
RT   "Characterization of rat liver mannan-binding protein gene.";
RL   J. Biochem. 111:66-73(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=86196130; PubMed=3009480;
RA   Drickamer K., Dordal M.S., Reynolds L.;
RT   "Mannose-binding proteins isolated from rat liver contain
RT   carbohydrate-recognition domains linked to collagenous tails. Complete
RT   primary structures and homology with pulmonary surfactant
RT   apoprotein.";
RL   J. Biol. Chem. 261:6878-6887(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Liver;
RX   MEDLINE=87194686; PubMed=3032924;
RA   Oka S., Itoh N., Kawasaki T., Yamashina I.;
RT   "Primary structure of rat liver mannan-binding protein deduced from
RT   its cDNA sequence.";
RL   J. Biochem. 101:135-144(1987).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 133-244.
RX   MEDLINE=96132792; PubMed=8557671;
RA   Ng K.K.-S., Drickamer K., Weis W.I.;
RT   "Structural analysis of monosaccharide recognition by rat liver
RT   mannose-binding protein.";
RL   J. Biol. Chem. 271:663-674(1996).
CC   -!- FUNCTION: Binds mannose and N-acetylglucosamine in a calcium-
CC       dependent manner. Is capable of host defense against pathogens, by
CC       activating the classical complement pathway independently of the
CC       antibody.
CC   -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST
```

```
CC       PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
CC       APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
CC       LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL
CC       SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE
CC       PROTEIN.
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14103; AAA41554.1; -.
DR   EMBL; X05023; CAA28687.1; -.
DR   PIR; A24791; LNRTMC.
DR   PDB; 1RDI; 08-MAR-96.
DR   PDB; 1RDJ; 08-MAR-96.
DR   PDB; 1RDK; 08-MAR-96.
DR   PDB; 1RDL; 08-MAR-96.
DR   PDB; 1RDM; 08-MAR-96.
DR   PDB; 1RDN; 08-MAR-96.
DR   PDB; 1RDO; 08-MAR-96.
DR   PDB; 1BV4; 29-DEC-99.
DR   PDB; 1KZA; 05-JUL-02.
DR   PDB; 1KZB; 05-JUL-02.
DR   PDB; 1KZC; 05-JUL-02.
DR   PDB; 1KZD; 05-JUL-02.
DR   PDB; 1KZE; 05-JUL-02.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Complement pathway; Membrane; Mannose-binding; Calcium; Repeat;
KW   Signal; Collagen; Lectin; Glycoprotein; Hydroxylation; 3D-structure.
FT   SIGNAL       1     18
FT   CHAIN       19    244       MANNOSE-BINDING PROTEIN C.
FT   DOMAIN      38     96       COLLAGEN-LIKE.
FT   DOMAIN     129    241       C-TYPE LECTIN.
FT   MOD_RES     43     43       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     58     58       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     69     69       HYDROXYLATION.
FT   MOD_RES     78     78       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     81     81       HYDROXYLATION (POTENTIAL).
FT   DISULFID    29     29       INTERCHAIN.
FT   DISULFID    34     34       INTERCHAIN.
FT   DISULFID   151    240
FT   DISULFID   218    232
FT   CONFLICT    38     39       GL -> AW (IN REF. 2).
FT   STRAND     134    137
FT   STRAND     142    142
FT   HELIX      144    153
FT   TURN       154    155
FT   STRAND     157    158
FT   HELIX      164    173
FT   STRAND     178    183
FT   TURN       188    189
FT   STRAND     192    193
FT   TURN       194    195
FT   STRAND     198    198
FT   STRAND     204    204
FT   TURN       206    207
FT   TURN       213    214
FT   STRAND     218    221
FT   TURN       223    224
FT   STRAND     227    230
FT   TURN       232    233
FT   STRAND     236    242
SQ   SEQUENCE   244 AA;  26014 MW;  F0706E2AA9331531 CRC64;

Alignment Scores:
Pred. No.:                9.47e-58        Length:          244
Score:                    785.00          Matches:         151
Percent Similarity:       75.20%          Conservative:    34
Best Local Similarity:    61.38%          Mismatches:      57
Query Match:              57.51%          Indels:          4
DB:                       1               Gaps:            3

US-10-054-536-2 (1-747) x MABC_RAT (1-244)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||   |||         |||||||||   ::::::   |||   |||:::
Db          1 MetSerLeuPheThrSer------PheLeuLeuLeuCysValLeuThrAlaValTyrAla 18

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||:::|||   |||   ||||||   :::||||||   ||||||||||||||||||
Db         19 GluThrLeuThr---GluGlyAlaGlnSerSerCysPro---ValIleAlaCysSerSer 36

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||:::|||||||||||||||||||||||||   ||||||   ||||||||||||||
Db         37 ProGlyLeuAsnGlyPheProGlyLysAspGlyHisAspGlyAlaLysGlyGluLysGly 56

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              |||||||||||||||||||||||||||||||||||||||||||:::||||||   |||
Db         57 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysValGlyProAlaGly 76

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
                ||||||      ||||||   |||      |||   |||||||||   |||:::|||
Db         77 ProProGlyAsnProGlySerLysGlyAlaThrGlyProLysGlyAspArgGlyGluSer 96

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
                 :::   |||::::::            |||      ||||||::::::|||:::
Db         97 ValGluPheAspThrThrAsnIleAspLeuGluIleAlaAlaLeuArgSerGluLeuArg 116

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
                 ::::::||||||:::      |||:::   :::   ||||||   |||:::|||:::
Db        117 AlaMetArgLysTrpValLeuLeuSerMetSerGluAsnValGlyLysLysTyrPheMet 136

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
                 ::::::         |||         :::   ||||||||||   :::   |||
Db        137 SerSerValArgArgMetProLeuAsnArgAlaLysAlaLeuCysSerGluLeuGlnGly 156

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
                 :::||||||||||||||||||||   ||||||   |||||||||||:::   |||:::
Db        157 ThrValAlaThrProArgAsnAlaGluGluAsnArgAlaIleGlnAsnValAlaLysAsp 176

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
                 |||||||||||||||||||||:::::|||||      |||   ||||||||||||
Db        177 ValAlaPheLeuGlyIleThrAspGlnArgThrGluAsnValPheGluAspLeuThrGly 196

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
                 ||||||:::   |||||||||||||||||||||||||||||||   ||||||   |||
Db        197 AsnArgValArgTyrThrAsnTrpAsnGluGlyGluProAsnAsnValGlySerGlyGlu 216

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
                 :::||||||:::||||||   ||||||:::||||||||||||||||||||   |||
Db        217 AsnCysValValLeuLeuThrAsnGlyLysTrpAsnAspValProCysSerAspSerPhe 236

Qy        721 CTGGCCGTCTGTGAGTTC 738
                 |||   ||||||||||||
Db        237 LeuValValCysGluPhe 242

RESULT 5
MABA_MOUSE
ID   MABA_MOUSE     STANDARD;      PRT;   239 AA.
AC   P39039;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Mannose-binding protein A precursor (MBP-A) (Mannan-binding protein)
```

```
DE   (RA-reactive factor polysaccharide-binding component P28B polypeptide)
DE   (RARF P28B).
GN   MBL1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBA/J; TISSUE=Liver;
RX   MEDLINE=91302823; PubMed=1712818;
RA   Sastry K., Zahedi K., Lelias J.M., Whitehead A.S., Ezekowitz R.A.;
RT   "Molecular characterization of the mouse mannose-binding proteins.
RT   The mannose-binding protein A but not C is an acute phase reactant.";
RL   J. Immunol. 147:692-697(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=92345256; PubMed=1637828;
RA   Kuge S., Ihara S., Watanabe E., Watanabe M., Takishima K., Suga T.,
RA   Mamiya G., Kawakami M.;
RT   "cDNAs and deduced amino acid sequences of subunits in the binding
RT   component of mouse bactericidal factor, Ra-reactive factor:
RT   similarity to mannose-binding proteins.";
RL   Biochemistry 31:6943-6950(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=95284466; PubMed=7766991;
RA   Sastry R., Wang J.S., Brown D.C., Ezekowitz R.A., Tauber A.I.,
RA   Sastry K.N.;
RT   "Characterization of murine mannose-binding protein genes Mbl1 and
RT   Mbl2 reveals features common to other collectin genes.";
RL   Mamm. Genome 6:103-110(1995).
CC   -!- FUNCTION: Binds mannose and N-acetylglucosamine in a calcium-
CC       dependent manner. Is capable of host defense against pathogens, by
CC       activating the classical complement pathway independently of the
CC       antibody.
CC   -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST
CC       PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
CC       APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
CC       LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL
CC       SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE
CC       PROTEIN.
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D11441; BAA02006.1; -.
DR   EMBL; S42292; AAB19342.1; -.
DR   EMBL; U09010; AAA82009.1; -.
DR   EMBL; U09007; AAA82009.1; JOINED.
DR   EMBL; U09008; AAA82009.1; JOINED.
DR   EMBL; U09009; AAA82009.1; JOINED.
DR   PIR; A46466; LNMSMA.
DR   HSSP; P19999; 1RTM.
DR   MGD; MGI:96923; Mbl1.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Complement pathway; Membrane; Mannose-binding; Calcium; Repeat;
KW   Signal; Collagen; Lectin; Glycoprotein; Hydroxylation.
FT   SIGNAL       1    17       BY SIMILARITY.
FT   CHAIN       18   239       MANNOSE-BINDING PROTEIN A.
FT   DOMAIN      37    89       COLLAGEN-LIKE.
FT   DOMAIN     144   239       C-TYPE LECTIN (SHORT FORM).
FT   DISULFID   146   235       BY SIMILARITY.
FT   DISULFID   213   227       BY SIMILARITY.
SQ   SEQUENCE   239 AA;  25396 MW;  D222F1A748D424D9 CRC64;

Alignment Scores:
Pred. No.:                1.26e-51        Length:          239
Score:                    711.50          Matches:         144
Percent Similarity:       69.48%          Conservative:    29
Best Local Similarity:    57.83%          Mismatches:      63
Query Match:              52.12%          Indels:          13
DB:                       1               Gaps:            4

US-10-054-536-2 (1-747) x MABA_MOUSE (1-239)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||   |||   |||   ||||||   :::||||||   :::|||:::   |||   |||
Db          1 MetLeuLeuLeuProLeuLeuPro---ValLeuLeuCysValValSerValSerSerSer 19

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
                :::   ||||||||||||     |||||||||   :::||||||||||||
Db         20 GlySerGlnThrCysGluAspThrLeuLysThrCys---SerValIleAlaCys------ 36

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
                                   |||:::|||||||||||||||   ||||||||||||||||
Db         37 --------------------GlyArgAspGlyArgAspGlyProLysGlyGluLysGly 49

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 69

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              :::   |||      |||||||||||||||||||||||||||||||   |||   :::
Db         70 SerValGlySerProGlySerProGlyProLysGlyGlnLysGlyAspHisGlyAspAsn 89

Qy        301 CCGGATGGTGATAGTAGCCTGGCT------GCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
                        :::      ||||||         :::|||   :::   |||:::::::::
Db         90 ArgAlaIleGluGluLysLeuAlaAsnMetGluAlaGluIleArgIleLeuLysSerLys 109

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
              :::           |||          ||||||:::||||||:::   |||   |||
Db        110 LeuGlnLeuThrAsnLysLeuHisAlaPheSerMetGlyLysLysSerGlyLysLysLeu 129

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              |||:::||||||   |||   |||   |||   |||||||||:::||||||   :::
Db        130 PheValThrAsnHisGluLysMetProPheSerLysValLysSerLeuCysThrGluLeu 149

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
              |||   :::||||||   ||||||||||||   ||||||   ||||||||   :::
Db        150 GlnGlyThrValAlaIleProArgAsnAlaGluGluAsnLysAlaIleGlnGluValAla 169

Qy        535 AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTG 594
                          ||||||||||||||||||||||||   ||||||||||||||:::   :::
Db        170 ThrGlyIleAlaPheLeuGlyIleThrAspGluAlaThrGluGlyGlnPheMetTyrVal 189

Qy        595 ACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCT 654
              ||||||   ||||||||||||:::||||||   :::   ||||||||||||   ||||||
Db        190 ThrGlyGlyArgLeuThrTyrSerAsnTrpLysLysAspGluProAsnAsnHisGlySer 209

Qy        655 GATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACC 714
                 ||||||||||||::::::|||   ||||||   |||||||||:::   |||
Db        210 GlyGluAspCysValIleIleLeuAspAsnGlyLeuTrpAsnAspIleSerCysGlnAla 229

Qy        715 TCCCATCTGGCCGTCTGTGAGTTCCCT 741
              |||      ||||||||||||||||||
Db        230 SerPheLysAlaValCysGluPhePro 238

RESULT 6
MABA_RAT
```

```
ID   MABA_RAT       STANDARD;      PRT;   238 AA.
AC   P19999;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Mannose-binding protein A precursor (MBP-A) (Mannan-binding protein).
GN   MBL1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=86196130; PubMed=3009480;
RA   Drickamer K., Dordal M.S., Reynolds L.;
RT   "Mannose-binding proteins isolated from rat liver contain
RT   carbohydrate-recognition domains linked to collagenous tails.
RT   Complete primary structures and homology with pulmonary surfactant
RT   apoprotein.";
RL   J. Biol. Chem. 261:6878-6887(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=87137502; PubMed=3029088;
RA   Drickamer K., McCreary V.;
RT   "Exon structure of a mannose-binding protein gene reflects its
RT   evolutionary relationship to the asialoglycoprotein receptor and
RT   nonfibrillar collagens.";
RL   J. Biol. Chem. 262:2582-2589(1987).
RN   [3]
RP   SEQUENCE OF 18-42.
RX   MEDLINE=87222358; PubMed=3584121;
RA   Ikeda K., Sannoh T., Kawasaki N., Kawasaki T., Yamashina I.;
RT   "Serum lectin with known structure activates complement through the
RT   classical pathway.";
RL   J. Biol. Chem. 262:7451-7454(1987).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 124-238.
RX   MEDLINE=92086855; PubMed=1721241;
RA   Weis W.I., Kahn R., Fourme R., Drickamer K., Hendrickson W.A.;
RT   "Structure of the calcium-dependent lectin domain from a rat mannose-
RT   binding protein determined by MAD phasing.";
RL   Science 254:1608-1615(1991).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 124-238.
RX   MEDLINE=93063338; PubMed=1436090;
RA   Weis W.I., Drickamer K., Hendrickson W.A.;
RT   "Structure of a C-type mannose-binding protein complexed with an
RT   oligosaccharide.";
RL   Nature 360:127-134(1992).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 84-238.
RX   MEDLINE=95219384; PubMed=7704532;
RA   Weis W.I., Drickamer K.;
RT   "Trimeric structure of a C-type mannose-binding protein.";
RL   Structure 2:1227-1240(1994).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 90-238.
RX   MEDLINE=99119227; PubMed=9922165;
RA   Ng K.K.-S., Park-Snyder S., Weis W.I.;
RT   "Ca2+-dependent structural changes in C-type mannose-binding
RT   proteins.";
RL   Biochemistry 37:17965-17976(1998).
CC   -!- FUNCTION: Binds mannose and N-acetylglucosamine in a calcium-
CC       dependent manner. Is capable of host defense against pathogens, by
CC       activating the classical complement pathway independently of the
CC       antibody.
CC   -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST
CC       PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
CC       APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
CC       LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL SPACE
CC       OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE PROTEIN.
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14105; AAA98781.1; -.
DR   EMBL; M14104; AAA98781.1; JOINED.
DR   PIR; B24791; LNRTMA.
DR   PDB; 1MSB; 15-JAN-92.
DR   PDB; 2MSB; 31-OCT-93.
DR   PDB; 1AFA; 03-APR-96.
DR   PDB; 1AFB; 03-APR-96.
DR   PDB; 1AFD; 03-APR-96.
DR   PDB; 1RTM; 07-FEB-95.
DR   PDB; 1KMB; 12-FEB-97.
DR   PDB; 2KMB; 12-FEB-97.
DR   PDB; 3KMB; 12-FEB-97.
DR   PDB; 4KMB; 12-FEB-97.
DR   PDB; 1YTT; 10-JUN-96.
DR   PDB; 1BCH; 17-JUN-98.
DR   PDB; 1BCJ; 17-JUN-98.
DR   PDB; 1BUU; 09-SEP-98.
DR   PDB; 1FIF; 10-JAN-01.
DR   PDB; 1FIH; 10-JAN-01.
DR   PDB; 1KWT; 05-JUL-02.
DR   PDB; 1KWU; 05-JUL-02.
DR   PDB; 1KWV; 05-JUL-02.
DR   PDB; 1KWW; 05-JUL-02.
DR   PDB; 1KWX; 05-JUL-02.
DR   PDB; 1KWY; 05-JUL-02.
DR   PDB; 1KWZ; 05-JUL-02.
DR   PDB; 1KX0; 05-JUL-02.
DR   PDB; 1KX1; 05-JUL-02.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Complement pathway; Membrane; Mannose-binding; Calcium; Repeat;
KW   Signal; Collagen; Lectin; Glycoprotein; Hydroxylation; 3D-structure.
FT   SIGNAL        1     17
FT   CHAIN        18    238       MANNOSE-BINDING PROTEIN A.
FT   DOMAIN       39     88       COLLAGEN-LIKE.
FT   DOMAIN      143    238       C-TYPE LECTIN (SHORT FORM).
FT   MOD_RES      43     43       HYDROXYLATION (POTENTIAL).
FT   MOD_RES      61     61       HYDROXYLATION.
FT   MOD_RES      67     67       HYDROXYLATION.
FT   MOD_RES      73     73       HYDROXYLATION.
FT   MOD_RES      78     78       HYDROXYLATION (POTENTIAL).
FT   DISULFID    145    234
FT   DISULFID    212    226
FT   CONFLICT    156    156       R -> K (IN REF. 2).
FT   HELIX        91    119
FT   TURN        120    121
FT   TURN        124    125
FT   STRAND      128    136
FT   HELIX       138    147
FT   TURN        148    149
FT   STRAND      151    152
FT   HELIX       158    168
FT   STRAND      172    177
FT   TURN        182    183
FT   STRAND      186    187
FT   TURN        188    189
```

```
FT   STRAND      192    192
FT   STRAND      198    198
FT   TURN        200    201
FT   TURN        207    208
FT   STRAND      212    215
FT   TURN        217    218
FT   STRAND      221    224
FT   TURN        226    227
FT   STRAND      230    236
SQ   SEQUENCE   238 AA;  25308 MW;  1A927482B8A8CB3D CRC64;

Alignment Scores:
Pred. No.:              3.98e-51      Length:          238
Score:                  705.50        Matches:         143
Percent Similarity:     70.08%        Conservative:    28
Best Local Similarity:  58.61%        Mismatches:      60
Query Match:            51.68%        Indels:          13
DB:                     1             Gaps:            4

US-10-054-536-2 (1-747) x MABA_RAT (1-238)

Qy         19 CTCCCTCTCCTT---CTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGT 75
              ||||||||||||   ||||||   :::|||:::   |||   |||   :::   ||||||
Db          4 LeuProLeuLeuValLeuLeuCysValValSerValSerSerSerGlySerGlnThrCys 23

Qy         76 GAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGC 135
              |||:::      |||||||||   :::||||||||||
Db         24 GluGluThrLeuLysThrCys---SerValIleAlaCys--------------------- 35

Qy        136 TTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGG 195
                 |||:::|||||||||||||||   ||||||||||||||||||||||||||||||
Db         36 ------GlyArgAspGlyArgAspGlyProLysGlyGluLysGlyGluProGlyGlnGly 53

Qy        196 CTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCT 255
              |||||||||||||||||||||||||||||||||||||||:::   |||
Db         54 LeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGlySerValGlyAlaPro 73

Qy        256 GGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGT 315
              ||||||   ||||||||||||||||||||||   |||   |||          :::
Db         74 GlySerGlnGlyProLysGlyGlnLysGlyAspArgGlyAspSerArgAlaIleGluVal 93

Qy        316 AGCCTGGCT------GCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAA 369
              ||||||         :::|||         |||:::::::::::
Db         94 LysLeuAlaAsnMetGluAlaGluIleAsnThrLeuLysSerLysLeuGluLeuThrAsn 113

Qy        370 AAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGT 429
              |||          ||||||:::|||||:::   |||   |||||||||:::||||||
Db        114 LysLeuHisAlaPheSerMetGlyLysLysSerGlyLysLysPhePheValThrAsnHis 133

Qy        430 GAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCC 489
              |||   |||   |||   |||||||||||||||||   :::   :::   :::||||||
Db        134 GluArgMetProPheSerLysValLysAlaLeuCysSerGluLeuArgGlyThrValAla 153

Qy        490 ACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTC 549
              ||||||||||||   ||||||   ||||||||   :::   |||      ||||||
Db        154 IleProArgAsnAlaGluGluAsnLysAlaIleGlnGluValAlaLysThrSerAlaPhe 173

Qy        550 CTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTG 609
              |||||||||||||||   |||||||||||||||:::   :::||||||   ||||||
Db        174 LeuGlyIleThrAspGluValThrGluGlyGlnPheMetTyrValThrGlyGlyArgLeu 193

Qy        610 ACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTA 669
              ||||||:::||||||   :::   ||||||||:::   ||||||   ||||||||||||
Db        194 ThrTyrSerAsnTrpLysLysAspGluProAsnAspHisGlySerGlyGluAspCysVal 213

Qy        670 TTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTC 729
              ::::::   ||||||   |||||||||:::   |||      ||||||   ||||||
Db        214 ThrIleValAspAsnGlyLeuTrpAsnAspIleSerCysGlnAlaSerHisThrAlaVal 233

Qy        730 TGTGAGTTCCCT 741
              ||||||||||||
Db        234 CysGluPhePro 237
```

```
RESULT 7
PSPD_HUMAN
ID   PSPD_HUMAN     STANDARD;      PRT;   375 AA.
AC   P35247;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN   SFTPD OR SFTP4 OR PSPD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=93155122; PubMed=8428971;
RA   Crouch E., Rust K., Veile R., Donis-Keller H., Grosso L.;
RT   "Genomic organization of human surfactant protein D (SP-D). SP-D is
RT   encoded on chromosome 10q22.2-23.1.";
RL   J. Biol. Chem. 268:2976-2983(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 214-243.
RC   TISSUE=Amniotic fluid, and Lung;
RX   MEDLINE=92322003; PubMed=1339284;
RA   Lu J., Willis A.C., Reid K.B.M.;
RT   "Purification, characterization and cDNA cloning of human lung
RT   surfactant protein D.";
RL   Biochem. J. 284:795-802(1992).
RN   [3]
RP   SEQUENCE OF 60-375 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Lung;
RX   MEDLINE=91378578; PubMed=1898081;
RA   Rust K., Grosso L., Zhang V., Chang D., Persson A., Longmore W.,
RA   Cai G.-Z., Crouch E.;
RT   "Human surfactant protein D: SP-D contains a C-type lectin
RT   carbohydrate recognition domain.";
RL   Arch. Biochem. Biophys. 290:116-126(1991).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=99197291; PubMed=10368295;
RA   Hakansson K., Lim N.K., Hoppe H.-J., Reid K.B.M.;
RT   "Crystal structure of the trimeric alpha-helical coiled-coil and the
RT   three lectin domains of human lung surfactant protein D.";
RL   Structure 7:255-264(1999).
CC   -!- FUNCTION: Contributes to the lung's defense against inhaled
CC       microorganisms. Binds strongly maltose residues and to a lesser
CC       extent other alpha-glucosyl moieties. It could participate in the
CC       extracellular reorganization or turnover of pulmonary surfactant.
CC   -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- PTM: The N-terminus is blocked.
CC   -!- MISCELLANEOUS: Pulmonary surfactant consists of 90% lipid and 10%
CC       protein. There are 4 surfactant-associated protein: 2 collagenous,
CC       carbohydrate-binding glycoproteins (SP-A and SP-D) and 2 small
CC       hydrophobic proteins (SP-B and SP-C).
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L05485; AAB59450.1; -.
DR   EMBL; L05483; AAB59450.1; JOINED.
DR   EMBL; L05484; AAB59450.1; JOINED.
DR   EMBL; X65018; CAA46152.1; -.
DR   PIR; A45225; A45225.
DR   PDB; 1B08; 29-NOV-99.
```

```
DR   PDB; 1M7L; 27-NOV-02.
DR   Genew; HGNC:10803; SFTPD.
DR   MIM; 178635; -.
DR   GO; GO:0005530; F:lectin; TAS.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 4.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW   Signal; Lectin; Collagen; Repeat; Coiled coil; 3D-structure.
FT   SIGNAL       1    20       BY SIMILARITY.
FT   CHAIN       21   375       PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT                              D.
FT   DOMAIN      46   222       COLLAGEN-LIKE.
FT   DOMAIN     223   252       COILED COIL (POTENTIAL).
FT   DOMAIN     279   375       C-TYPE LECTIN (SHORT FORM).
FT   DISULFID   281   373
FT   DISULFID   351   365
FT   CARBOHYD    90    90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MOD_RES     78    78       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     87    87       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     96    96       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     99    99       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    171   171       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    177   177       HYDROXYLATION (BY SIMILARITY).
FT   CONFLICT    31    31       M -> T (IN REF. 2).
FT   CONFLICT    59    59       P -> F (IN REF. 3).
FT   CONFLICT   122   122       A -> P (IN REF. 2).
FT   CONFLICT   180   180       T -> A (IN REF. 2).
FT   CONFLICT   206   206       D -> P (IN REF. 3).
FT   CONFLICT   374   374       E -> EH (IN REF. 3).
FT   HELIX      227   253
FT   TURN       254   256
FT   STRAND     257   260
FT   TURN       261   262
FT   STRAND     263   272
FT   HELIX      274   283
FT   TURN       284   285
FT   STRAND     287   288
FT   HELIX      294   307
FT   TURN       308   308
FT   STRAND     311   316
FT   TURN       321   322
FT   STRAND     325   325
FT   TURN       327   328
FT   STRAND     331   331
FT   STRAND     337   337
FT   TURN       339   340
FT   HELIX      345   347
FT   STRAND     351   354
FT   TURN       356   357
FT   STRAND     360   363
FT   TURN       365   366
FT   STRAND     369   375
SQ   SEQUENCE   375 AA;  37702 MW;  2986B2699FC01A6A CRC64;

Alignment Scores:
Pred. No.:                4.01e-30      Length:          375
Score:                    454.00        Matches:         99
Percent Similarity:       55.07%        Conservative:    26
Best Local Similarity:    43.61%        Mismatches:      82
Query Match:              33.26%        Indels:          20
DB:                       1             Gaps:            5

US-10-054-536-2 (1-747) x PSPD_HUMAN (1-375)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||:::   |||      |||      |||      |||   |||||||||:::

Db        149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy        178 GGGGAACCAGGCCAA----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   |||||:::          |||      |||      |||      |||

Db        169 GlyValProGlyGluArgGlyValProGlyAsnThrGlyAlaAlaGlySerAlaGlyAla 188

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||:::||||||      |||   ||||||   ||||||   ||||||

Db        189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy        286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
                ||||||   |||         |||:::|||   |||         ||||||   |||

Db        209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy        337 AAAGCT------------------------------CTGCAAACAGAAATGGCACGTATC 366
              :::                              ||||||        ::::::

Db        229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||      |||      |||:::   ||||||   |||   |||   |||

Db        249 LysLysValGluLeuProProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::      |||   :::   :::   ||||||   :::           :::

Db        269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
              |||:::||||||:::||||||||||||   |||:::|||   |||:::

Db        289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              |||   |||||||||   :::||||||   |||||||||||:::|||          |||

Db        309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
              |||      |||   |||:::||||||      ||||||||||||:::   |||

Db        329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySer 348

Qy        658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              ||||||||||||   :::      ||||||:::|||||||||      |||

Db        349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy        718 CATCTGGCCGTCTGTGAGTTC 738
                  |||   ||||||||||||

Db        369 ArgLeuValValCysGluPhe 375

RESULT 8
CL46_BOVIN
ID   CL46_BOVIN     STANDARD;      PRT;   371 AA.
AC   Q8MHZ9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Collectin-46 precursor (CL-46) (46 kDa collectin).
GN   CL46.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hansen S., Holm D., Moeller V., Vitved L., Bendixen C., Reid K.B.M.,
RA   Skjoedt K., Holmskov U.;
RT   "CL-46, a novel collectin highly expressed in the bovine thymus and
RT   liver.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Highly expressed in thymus and liver.
CC   -!- PTM: Hydroxylated (Potential).
CC   -!- SIMILARITY: Contains 1 collagenous domain.
```

```
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF509589; AAM34742.1; -.
DR   EMBL; AF509590; AAM34743.1; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;
KW   Collagen; Repeat; Calcium; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    371       COLLECTIN-46.
FT   DOMAIN       46    216       COLLAGEN-LIKE.
FT   DOMAIN      273    371       C-TYPE LECTIN (SHORT FORM).
FT   SITE        201    203       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    275    369       BY SIMILARITY.
FT   DISULFID    347    361       BY SIMILARITY.
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   371 AA;  37445 MW;  108AC45A91420E83 CRC64;

Alignment Scores:
Pred. No.:              9.5e-30      Length:        371
Score:                  449.50       Matches:       116
Percent Similarity:     41.82%       Conservative:   40
Best Local Similarity:  31.10%       Mismatches:     88
Query Match:            32.93%       Indels:        129
DB:                     1            Gaps:           12

US-10-054-536-2 (1-747) x CL46_BOVIN (1-371)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||   |||   |||          ||||||||||||:::            |||
Db          1 MetLeuLeuLeuProLeuSerValLeuLeuLeuLeuThrGlnProTrpArgSerLeuGly 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACC-----------TGCCCTGCAGTGATT 108
                 :::          :::|||||||||            |||      |||:::
Db         21 AlaGluMetLysIleTyr---SerGlnLysThrLeuAlaAsnGlyCysThrLeuValVal 39

Qy        109 GCCTGTAGCTCTCCA------GGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGC 162
                |||      |||      |||:::   |||      |||:::||||||||||:::|||
Db         40 ---CysArgProProGluGlyGlyLeuProGlyArgAspGlyGlnAspGlyArgGluGly 58

Qy        163 ACCAAGGGAGAAAAGGGGGAACCAGGCCAA---------------------------- 192
                :::||||||||||||:::||||||
Db         59 ProGlnGlyGluLysGlyAspProGlySerProGlyProAlaGlyArgAlaGlyArgPro 78

Qy        193 ----------------------------------------------------GGGCTC 198
                                                                   |||
Db         79 GlyProAlaGlyProIleGlyProLysGlyAspAsnGlySerAlaGlyGluProGlyPro 98

Qy        199 AGAGGCTTACAGGGCCCCCCT---------------------------------------- 219
                :::|||      ||||||||
Db         99 LysGlyAspThrGlyProProGlyProProGlyMetProGlyProAlaGlyArgGluGly 118

Qy        220 --------------GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCT------------ 252
                           |||   :::||||||||||||   |||||||||
Db        119 ProSerGlyLysGlnGlySerMetGlyProProGlyThrProGlyProLysGlyAspThr 138

Qy        253 --------------------------------TCTGGGTCACCAGGACCA--------- 270
                                             ||||||||||||||||
Db        139 GlyProLysGlyGlyMetGlyAlaProGlyMetGlnGlySerProGlyProAlaGlyLeu 158

Qy        270 ---------------------------------------------------------- 270

Db        159 LysGlyGluArgGlyAlaProGlyGluLeuGlyAlaProGlySerAlaGlyValAlaGly 178

Qy        271 ----------------------------------------------------------AAG 273
                                                                         |||
Db        179 ProAlaGlyAlaIleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLys 198

Qy        274 GGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTGGCT------ 324
                |||   :::||||||||||||   :::         |||:::|||   ||||||
Db        199 GlyAspArgGlyAspProGlyGluArgGlyAlaLysGlyGluSerGlyLeuAlaAspVal 218

Qy        325 ----------------------------GCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
                                                 :::   :::   ||||||
Db        219 AsnAlaLeuLysGlnArgValThrIleLeuGluGlyGlnLeuGlnArgLeuGlnAsnAla 238

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
                 :::|||   ||||||   :::   |||      |||:::   ||||||   |||
Db        239 PheSerArgTyrLysLysAlaValLeuPheProAspGlyGlnAlaValGlyLysLysIle 258

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
                |||   |||   |||   :::   ::::::         :::   ||||||   :::
Db        259 PheLysThrAlaGlyAlaValLysSerTyrSerAspAlaGlnGlnLeuCysArgGluAla 278

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
                :::      :::|||:::||||||:::||||||||||||   |||:::      |||:::
Db        279 LysGlyGlnLeuAlaSerProArgSerAlaAlaGluAsnGluAlaValAlaGlnLeuVal 298

Qy        535 AAG---------GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTT 585
                :::            :::|||||||||   :::   |||      |||||||||:::|||
Db        299 ArgAlaLysAsnAsnAspAlaPheLeuSerMetAsnAspIleSerThrGluGlyLysPhe 318

Qy        586 GTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCC------ 639
                  ||||||      |||   |||:::||||||      |||||||||
Db        319 ThrTyrProThrGlyGluSerLeuValTyrSerAsnTrpAlaSerGlyGluProAsnAsn 338

Qy        640 AACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGAC 699
                ||||||||||      |||:::||||||   :::   :::   |||:::|||||||||
Db        339 AsnAsnAlaGlyGlnProGluAsnCysValGlnIleTyrArgGluGlyLysTrpAsnAsp 358

Qy        700 GTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
                ||||||||||||         |||   :::|||||||||
Db        359 ValProCysSerGluProLeuLeuValIleCysGluPhe 371

RESULT 9
CONG_BOVIN
ID   CONG_BOVIN     STANDARD;      PRT;   371 AA.
AC   P23805; O97748;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Conglutinin precursor.
GN   CGN1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93213261; PubMed=8460993;
RA   Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT   "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL   Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93277452; PubMed=7684896;
RA   Lu J., Laursen S.B., Thiel S., Jensenius J.C., Reid B.M.;
RT   "The cDNA cloning of conglutinin and identification of liver as a
RT   primary site of synthesis of conglutinin in members of the Bovidae.";
RL   Biochem. J. 292:157-162(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=Liver;
RX   MEDLINE=94215917; PubMed=8163202;
RA   Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
RA   Tauber A.I., Sastry K.N.;
RT   "Bovine conglutinin (BC) mRNA expressed in liver: cloning and
RT   characterization of the BC cDNA reveals strong homology to surfactant
RT   protein-D.";
RL   Gene 141:277-281(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Semen;
RX   MEDLINE=94267222; PubMed=8207234;
RA   Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
RA   Tauber A.I., Sastry K.N.;
RT   "Bovine conglutinin gene exon structure reveals its evolutionary
RT   relationship to surfactant protein-D.";
RL   J. Immunol. 153:173-180(1994).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94128104; PubMed=8297370;
RA   Kawasaki N., Itoh N., Kawasaki T.;
RT   "Gene organization and 5'-flanking region sequence of conglutinin: a
RT   C-type mammalian lectin containing a collagen-like domain.";
RL   Biochem. Biophys. Res. Commun. 198:597-604(1994).
RN   [6]
RP   SEQUENCE OF 21-371.
RX   MEDLINE=91131556; PubMed=1993651;
RA   Lee Y.-M., Leiby K.R., Allar J., Paris K., Lerch B., Okarma T.B.;
RT   "Primary structure of bovine conglutinin, a member of the C-type
RT   animal lectin family.";
RL   J. Biol. Chem. 266:2715-2723(1991).
RN   [7]
RP   PRELIMINARY SEQUENCE OF 21-52.
RX   MEDLINE=87184551; PubMed=3566740;
RA   Young N.M., Leon M.A.;
RT   "The carbohydrate specificity of conglutinin and its homology to
RT   proteins in the hepatic lectin family.";
RL   Biochem. Biophys. Res. Commun. 143:645-651(1987).
CC   -!- FUNCTION: Calcium-dependent lectin-like protein which binds to a
CC       yeast cell wall extract and immune complexes through the
CC       complement component (C3bi). It is capable of binding nonreducing
CC       terminal N-acetylglucosamine, mannose, and fucose residues.
CC   -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D14085; BAA03170.1; -.
DR   EMBL; X71774; CAA50665.1; -.
DR   EMBL; L18871; AAA20126.1; -.
DR   EMBL; U06860; AAB60624.1; -.
DR   EMBL; U06854; AAB60624.1; JOINED.
DR   EMBL; U06855; AAB60624.1; JOINED.
DR   EMBL; U06856; AAB60624.1; JOINED.
DR   EMBL; U06857; AAB60624.1; JOINED.
DR   EMBL; U06858; AAB60624.1; JOINED.
DR   EMBL; U06859; AAB60624.1; JOINED.
DR   EMBL; D25302; BAA04983.2; -.
DR   EMBL; D25296; BAA04983.2; JOINED.
DR   EMBL; D25297; BAA04983.2; JOINED.
DR   EMBL; D25298; BAA04983.2; JOINED.
DR   EMBL; D25299; BAA04983.2; JOINED.
DR   EMBL; D25300; BAA04983.2; JOINED.
DR   EMBL; D25301; BAA04983.2; JOINED.
DR   PIR; I45878; I45878.
DR   PIR; JN0450; JN0450.
DR   HSSP; P35247; 1B08.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;
KW   Collagen; Repeat; Calcium; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    371       CONGLUTININ.
FT   DOMAIN       46    216       COLLAGEN-LIKE.
FT   DOMAIN      273    371       C-TYPE LECTIN (SHORT FORM).
FT   MOD_RES      63     63       HYDROXYLATION.
FT   MOD_RES      87     87       HYDROXYLATION.
FT   MOD_RES      99     99       HYDROXYLATION.
FT   MOD_RES     135    135       HYDROXYLATION.
FT   MOD_RES     141    141       HYDROXYLATION.
FT   MOD_RES     159    159       HYDROXYLATION.
FT   MOD_RES     162    162       HYDROXYLATION.
FT   MOD_RES     198    198       HYDROXYLATION.
FT   SITE        201    203       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    275    369       BY SIMILARITY.
FT   DISULFID    347    361       BY SIMILARITY.
FT   CARBOHYD    337    337       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    173    173       R -> H (IN REF. 2 AND 3).
FT   CONFLICT    210    210       K -> S (IN REF. 6).
FT   CONFLICT    218    218       V -> A (IN REF. 2).
FT   CONFLICT    272    272       E -> V (IN REF. 2).
SQ   SEQUENCE   371 AA;  37994 MW;  867BB41992544B1F CRC64;

Alignment Scores:
Pred. No.:                4.41e-29        Length:          371
Score:                    441.50          Matches:         91
Percent Similarity:       56.77%          Conservative:    39
Best Local Similarity:    39.74%          Mismatches:      74
Query Match:              32.34%          Indels:          25
DB:                       1               Gaps:            6

US-10-054-536-2 (1-747) x CONG_BOVIN (1-371)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::|||||||||   ||||||||||||                |||   ||||||||||||
Db        146 AlaProGlyIleGlnGlyPheProGlyPro---------SerGlyLeuLysGlyGluLys 162

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   |||||:::             |||      |||:::   ||||||   |||
Db        163 GlyAlaProGlyGluThrGlyAlaProGlyArgAlaGlyValThrGlyProSerGlyAla 182

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||      |||      |||   ||||||   ||||||   :::|||
Db        183 IleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGly 202

Qy        286 GACCCTGGAAAAAGT---CCGGATGGTGATAGTAGCCTGGCTGCC-------------- 327
              ||||||||:::::      |||:::|||   ||||||
Db        203 AspProGlyGluThrGlyAlaLysGlyGluSerGlyLeuAlaGluValAsnAlaLeuLys 222

Qy        328 --------------------TCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
                                  :::      |||        ::::::
Db        223 GlnArgValThrIleLeuAspGlyHisLeuArgArgPheGlnAsnAlaPheSerGlnTyr 242

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||   :::   |||      |||:::   ||||||   |||   |||   |||
Db        243 LysLysAlaValLeuPheProAspGlyGlnAlaValGlyGluLysIlePheLysThrAla 262

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::   ::::::         :::   ||||||   :::   :::      :::
Db        263 GlyAlaValLysSerTyrSerAspAlaGluGlnLeuCysArgGluAlaLysGlyGlnLeu 282
```

```
Qy        487 GCCACCCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG--------- 537
              |||:::||||||::::::|||||||||   |||:::      :::::::::
Db        283 AlaSerProArgSerSerAlaGluAsnGluAlaValThrGlnMetValArgAlaGlnGlu 302

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              :::   |||:::|||   :::   |||      |||||||||:::|||         |||
Db        303 LysAsnAlaTyrLeuSerMetAsnAspIleSerThrGluGlyArgPheThrTyrProThr 322

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT------GGT 651
              |||      |||   |||:::||||||   :::|||||||||||||||:::      |||
Db        323 GlyGluIleLeuValTyrSerAsnTrpAlaAspGlyGluProAsnAsnSerAspGluGly 342

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
                    |||:::||||||   :::      :::|||:::|||||||||||||||||||||
Db        343 GlnProGluAsnCysValGluIlePheProAspGlyLysTrpAsnAspValProCysSer 362

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
                    |||   :::|||||||||
Db        363 LysGlnLeuLeuValIleCysGluPhe 371
```

RESULT 10
PSPD_RAT
ID PSPD_RAT STANDARD; PRT; 374 AA.
AC P35248;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D)
DE (CP4).
GN SFTPD OR SFTP4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-33.
RC TISSUE=Lung;
RX MEDLINE=92112913; PubMed=1370483;
RA Shimizu H., Fisher J.H., Papst P., Benson B., Lau K., Mason R.J.,
RA Voelker D.R.;
RT "Primary structure of rat pulmonary surfactant protein D. cDNA and
RT deduced amino acid sequence.";
RL J. Biol. Chem. 267:1853-1857(1992).
RN [2]
RP SEQUENCE OF 73-95 AND 153-180.
RC TISSUE=Lung;
RX MEDLINE=90001186; PubMed=2675969;
RA Persson A., Chang D., Rust K., Moxley M., Longmore W., Crouch E.;
RT "Purification and biochemical characterization of CP4 (SP-D), a
RT collagenous surfactant-associated protein.";
RL Biochemistry 28:6361-6367(1989).
CC -!- FUNCTION: Contributes to the lung's defense against inhaled
CC   microorganisms. Binds strongly maltose residues and to a lesser
CC   extent other alpha-glucosyl moieties. It could participate in the
CC   extracellular reorganization or turnover of pulmonary surfactant.
CC -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: Pulmonary surfactant consists of 90% lipid and 10%
CC   protein. There are 4 surfactant-associated protein: 2 collagenous,
CC   carbohydrate-binding glycoproteins (SP-A and SP-D) and 2 small
CC   hydrophobic proteins (SP-B and SP-C).
CC -!- SIMILARITY: Contains 1 collagenous domain.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M81231; AAA42170.1; -.
DR PIR; A42046; A42046.
DR HSSP; P35247; 1B08.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat; Coiled coil.
FT SIGNAL   1   19
FT CHAIN   20   374   PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT   D.
FT DOMAIN   45   221   COLLAGEN-LIKE.
FT DOMAIN   222   253   COILED COIL (POTENTIAL).
FT DOMAIN   278   374   C-TYPE LECTIN (SHORT FORM).
FT DISULFID 280   372   BY SIMILARITY.
FT DISULFID 350   364   BY SIMILARITY.
FT CARBOHYD   89   89   N-LINKED (GLCNAC. . .).
FT MOD_RES   77   77   HYDROXYLATION.
FT MOD_RES   86   86   HYDROXYLATION.
FT MOD_RES   95   95   HYDROXYLATION.
FT MOD_RES   98   98   HYDROXYLATION.
FT MOD_RES   170   170   HYDROXYLATION.
FT MOD_RES   176   176   HYDROXYLATION.
FT CONFLICT   89   89   N -> E (IN REF. 2).
FT CONFLICT 164   164   K -> C (IN REF. 2).
SQ SEQUENCE 374 AA; 37561 MW; DB2BB5E399DB4A3C CRC64;

Alignment Scores:

Pred. No.:	4.86e-29	Length:	374
Score:	441.00	Matches:	110
Percent Similarity:	38.81%	Conservative:	34
Best Local Similarity:	29.65%	Mismatches:	99
Query Match:	32.31%	Indels:	128
DB:	1	Gaps:	7

US-10-054-536-2 (1-747) x PSPD_RAT (1-374)

```
Qy         10 TTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTG 69
              |||   |||:::   :::||||||:::      :::            :::|||
Db          4 PheLeuSerMetLeuValLeuLeuValGlnProLeuGlyAspLeuGlyAlaGluMetLys 23

Qy         70 ACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATC 129
              |||                    |||           :::         ||||||
Db         24 ThrLeuSerGlnArgSerIleThrAsnThrCysThrLeuValLeuCysSerProThrGlu 43

Qy        130 AACGGCTTCCCAGGCAAAGATGGGCGTGAT---------GGCACCAAGGGAGAAAAGGGG 180
              ||||||   ||||||:::|||||||||||         |||   :::|||||||||||
Db         44 AsnGlyLeuProGlyArgAspGlyArgAspGlyArgGluGlyProArgGlyGluLysGly 63

Qy        181 GAACCAGGC--------------------------------------------------- 189
              :::||||||
Db         64 AspProGlyLeuProGlyProMetGlyLeuSerGlyLeuProGlyProArgGlyProVal 83

Qy        190 ----------------------------------------CAAGGGCTCAGAGGCTTA 207
                                                     :::|||   |||||||||
Db         84 GlyProLysGlyGluAsnGlySerAlaGlyGluProGlyProLysGlyGluArgGlyLeu 103

Qy        208 CAGGGCCCCCCT------------------------------------------------ 219
                 |||||||||
Db        104 ValGlyProProGlySerProGlyIleSerGlyProAlaGlyLysGluGlyProSerGly 123

Qy        220 ------GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAG 273
                    |||   :::||||||   |||   |||||||||   |||      |||||||||
Db        124 LysGlnGlyAsnIleGlyProGlnGlyLysProGlyProLysGlyGluAlaGlyProLys 143

Qy        274 GGCCAAAAAGGAGACCCTGGA----------------------AAAAGTCCGGATGGT--- 309
              |||:::   |||   ||||||                      |||   |||   |||
Db        144 GlyGluValGlyAlaProGlyMetGlnGlySerAlaGlyAlaLysGlyProAlaGlyPro 163

Qy        309 ------------------------------------------------------------ 309
```

```
Db        164 LysGlyGluArgGlyAlaProGlyGluGlnGlyAlaProGlyAsnAlaGlyAlaAlaGly 183

Qy        309 ------------------------------------------------------------ 309

Db        184 ProAlaGlyProAlaGlyProGlnGlyAlaProGlySerArgGlyProProGlyLeuLys 203

Qy        310 ------------------------------------------------------GATAGT 315
                                                                     ||||||
Db        204 GlyAspArgGlyAlaProGlyAspArgGlyIleLysGlyGluSerGlyLeuProAspSer 223

Qy        316 AGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC--------- 366
              :::             :::   :::||||||      ::::::   |||:::
Db        224 AlaAlaLeuArgGlnGlnMetGluAlaLeuAsnGlyLysLeuGlnArgLeuGluAlaAla 243

Qy        367 ------------AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
                         ||||||         |||      |||:::   ||||||:::|||
Db        244 PheSerArgTyrLysLysAlaAlaLeuPheProAspGlyGlnSerValGlyAspLysIle 263

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              |||                          ||||||      |||   :::|||   :::
Db        264 PheArgAlaAlaAsnSerGluGluProPheGluAspAlaLysGluMetCysArgGlnAla 283

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
                    :::|||:::||||||:::|||   ||||||   |||:::|||   |||:::
Db        284 GlyGlyGlnLeuAlaSerProArgSerAlaThrGluAsnAlaAlaValGlnGlnLeuVal 303

Qy        535 ---------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTT 585
                        :::   |||||||||   :::||||||      |||||||||:::|||
Db        304 ThrAlaHisSerLysAlaAlaPheLeuSerMetThrAspValGlyThrGluGlyLysPhe 323

Qy        586 GTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAAT 645
                    ||||||      |||   |||:::||||||      ||||||||||||||||
Db        324 ThrTyrProThrGlyGluAlaLeuValTyrSerAsnTrpAlaProGlyGluProAsnAsn 343

Qy        646 GCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCC 705
                  |||      |||:::||||||   :::      ||||||||||||||||
Db        344 AsnGlyGlyAlaGluAsnCysValGluIlePheThrAsnGlyGlnTrpAsnAspLysAla 363

Qy        706 TGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
                  |||            |||   :::|||||||||
Db        364 CysGlyGluGlnArgLeuValIleCysGluPhe 374
```

```
RESULT 11
PSPD_MOUSE
ID   PSPD_MOUSE     STANDARD;      PRT;   374 AA.
AC   P50404;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN   SFTPD OR SFTP4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX   MEDLINE=96094460; PubMed=7499852;
RA   Motwani M., White R.A., Guo N., Dowler L.L., Tauber A.I., Sastry K.N.;
RT   "Mouse surfactant protein-D. cDNA cloning, characterization, and gene
RT   localization to chromosome 14.";
RL   J. Immunol. 155:5671-5677(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=99244602; PubMed=10226065;
RA   Lawson P.R., Perkins V.C., Holmskov U., Reid K.B.;
RT   "Genomic organization of the mouse gene for lung surfactant protein
RT   D.";
RL   Am. J. Respir. Cell Mol. Biol. 20:953-963(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Fisher J.H., Sheftelyevich V.V.;
RT   "Surfactant protein-D regulates surfactant phospholipid homeostasis in
RT   vivo.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Contributes to the lung's defense against inhaled
CC       microorganisms. Binds strongly maltose residues and to a lesser
CC       extent other alpha-glucosyl moieties. It could participate in the
CC       extracellular reorganization or turnover of pulmonary surfactant.
CC   -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- MISCELLANEOUS: Pulmonary surfactant consists of 90% lipid and 10%
CC       protein. There are 4 surfactant-associated protein: 2 collagenous,
CC       carbohydrate-binding glycoproteins (SP-A and SP-D) and 2 small
CC       hydrophobic proteins (SP-B and SP-C).
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L40156; AAA92021.1; -.
DR   EMBL; AF047742; AAD31380.1; -.
DR   EMBL; AF047741; AAD31380.1; JOINED.
DR   EMBL; AF192134; AAF15277.1; -.
DR   EMBL; BC003705; AAH03705.1; -.
DR   HSSP; P35247; 1B08.
DR   MGD; MGI:109515; Sftpd.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW   Signal; Lectin; Collagen; Repeat; Coiled coil.
FT   SIGNAL        1     19       BY SIMILARITY.
FT   CHAIN        20    374       PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT                                 D.
FT   DOMAIN       45    221       COLLAGEN-LIKE.
FT   DOMAIN      222    253       COILED COIL (POTENTIAL).
FT   DOMAIN      278    374       C-TYPE LECTIN (SHORT FORM).
FT   DISULFID    280    372       BY SIMILARITY.
FT   DISULFID    350    364       BY SIMILARITY.
```

```
FT   CARBOHYD     89      89       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   374 AA;  37688 MW;  FE034261263F43E4 CRC64;

Alignment Scores:
Pred. No.:                7.14e-29        Length:          374
Score:                    439.00          Matches:         110
Percent Similarity:       39.89%          Conservative:    40
Best Local Similarity:    29.26%          Mismatches:      92
Query Match:              32.16%          Indels:          134
DB:                       1               Gaps:            9

US-10-054-536-2 (1-747) x PSPD_MOUSE (1-374)

Qy          7 CTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTG-----------GCAGCGTCT 54
              :::   |||   |||   :::|||:::|||   :::                  :::
Db          1 MetLeuProPheLeuSerMetLeuValLeuLeuValGlnProLeuGlyAsnLeuGlyAla 20

Qy         55 TACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGT 114
                  ::::::::::::                  ||||||      |||:::
Db         21 GluMetLysSerLeuSerGlnArgSerValProAsnThrCysThrLeuValMetCys--- 39

Qy        115 AGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGAT---------GGCACC 165
                ||||||      ||||||   ||||||:::||||||||||||         |||
Db         40 ---SerProThrGluAsnGlyLeuProGlyArgAspGlyArgAspGlyArgGluGlyPro 58

Qy        166 AAGGGAGAAAAGGGGGAACCAGGC---------------------------------- 189
              :::|||||||||||:::||||||
Db         59 ArgGlyGluLysGlyAspProGlyLeuProGlyProMetGlyLeuSerGlyLeuGlnGly 78

Qy        190 ----------------------------------------------------------CAA 192
                                                                       :::
Db         79 ProThrGlyProValGlyProLysGlyGluAsnGlySerAlaGlyGluProGlyProLys 98

Qy        193 GGGCTCAGAGGCTTACAGGGCCCCCCT---------------------------------- 219
              |||   ||||||||   |||||||||
Db         99 GlyGluArgGlyLeuSerGlyProProGlyLeuProGlyIleProGlyProAlaGlyLys 118

Qy        220 --------------------GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGG 258
                                 |||   :::||||||   |||   |||||||||   |||
Db        119 GluGlyProSerGlyLysGlnGlyAsnIleGlyProGlnGlyLysProGlyProLysGly 138

Qy        259 TCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA----------------------- 294
                    ||||||||||||:::   |||   ||||||
Db        139 GluAlaGlyProLysGlyGluValGlyAlaProGlyMetGlnGlySerThrGlyAlaLys 158

Qy        294 ------------------------------------------------------------ 294

Db        159 GlySerThrGlyProLysGlyGluArgGlyAlaProGlyValGlnGlyAlaProGlyAsn 178

Qy        295 ----------------------------------------------------------AAAAGT 300
                                                                         :::
Db        179 AlaGlyAlaAlaGlyProAlaGlyProAlaGlyProGlnGlyAlaProGlySerArgGly 198

Qy        301 CCGGATGGT------------------------------------------------- 309
              |||   |||
Db        199 ProProGlyLeuLysGlyAspArgGlyValProGlyAspArgGlyIleLysGlyGluSer 218

Qy        310 ---------GATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
                      ||||||:::             :::   :::||||||:::   ::::::
Db        219 GlyLeuProAspSerAlaAlaLeuArgGlnGlnMetGluAlaLeuLysGlyLysLeuGln 238

Qy        361 CGTATC--------------------AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAA 399
              |||:::                    :::|||         |||      |||:::
Db        239 ArgLeuGluValAlaPheSerHisTyrGlnLysAlaAlaLeuPheProAspGlyArgSer 258

Qy        400 GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCC 459
              ||||||:::|||   |||   |||                ||||||      :::
Db        259 ValGlyAspLysIlePheArgThrAlaAspSerGluLysProPheGluAspAlaGlnGlu 278

Qy        460 TTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCC 519
              :::|||   :::            :::|||:::||||||:::|||   ||||||   |||
Db        279 MetCysLysGlnAlaGlyGlyGlnLeuAlaSerProArgSerAlaThrGluAsnAlaAla 298

Qy        520 ATTCAGAATCTCATC---------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAG 570
              ||||||   ||||||           :::   |||||||||   :::||||||
Db        299 IleGlnGlnLeuIleThrAlaHisAsnLysAlaAlaPheLeuSerMetThrAspValGly 318

Qy        571 ACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAG 630
              |||||||||:::|||         ||||||      |||   |||:::||||||
Db        319 ThrGluGlyLysPheThrTyrProThrGlyGluProLeuValTyrSerAsnTrpAlaPro 338

Qy        631 GGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAG 690
              ||||||||||||||   |||      |||:::||||||   :::      |||||||||
Db        339 GlyGluProAsnAsnAsnGlyGlyAlaGluAsnCysValGluIlePheThrAsnGlyGln 358

Qy        691 TGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
              |||||||||      |||           |||   :::|||||||||
Db        359 TrpAsnAspLysAlaCysGlyGluGlnArgLeuValIleCysGluPhe 374
```

```
RESULT 12
CL43_BOVIN
ID   CL43_BOVIN     STANDARD;      PRT;   321 AA.
AC   P42916; Q8WMF4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collectin-43 precursor (CL-43) (43 kDa collectin).
GN   CL43.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=22414671; PubMed=12527419;
RA   Hansen S., Holm D., Moeller V., Vitved L., Bendixen C., Skjoedt K.,
RA   Holmskov U.;
RT   "Genomic and molecular characterization of CL-43 and its proximal
RT   promoter.";
RL   Biochim. Biophys. Acta 1625:1-10(2003).
RN   [2]
RP   SEQUENCE OF 21-321 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=94216283; PubMed=8163480;
RA   Lim B.-L., Willis A.C., Reid K.B.M., Lu J., Laursen S.B.,
RA   Jensenius J.C., Holmskov U.;
RT   "Primary structure of bovine collectin-43 (CL-43). Comparison with
RT   conglutinin and lung surfactant protein-D.";
RL   J. Biol. Chem. 269:11820-11824(1994).
CC   -!- FUNCTION: Lectin that binds to various sugars: mannose = manNAc >
CC       fucose > glcNAc > glucose = maltose > galactose > lactose >
CC       galNAc. Could play a role in immune defense.
CC   -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Liver specific.
CC   -!- PTM: Hydroxylated (Potential).
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AY071821; AAL61855.1; -.
DR   EMBL; AY071822; AAL61856.1; -.
DR   EMBL; X75912; CAA53511.1; ALT_SEQ.
DR   HSSP; P35247; 1B08.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
```

```
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Collagen;
KW   Repeat; Calcium; Signal.
FT   SIGNAL        1    20
FT   CHAIN        21   321       COLLECTIN-43.
FT   DOMAIN       49   162       COLLAGEN-LIKE.
FT   DOMAIN      222   321       C-TYPE LECTIN (SHORT FORM).
FT   DISULFID    224   319       BY SIMILARITY.
FT   DISULFID    297   311       BY SIMILARITY.
FT   CONFLICT    125   125       T -> A (IN REF. 2).
FT   CONFLICT    286   286       N -> G (IN REF. 2).
SQ   SEQUENCE   321 AA;  33615 MW;  12BF120BB48861A1 CRC64;

Alignment Scores:
Pred. No.:              6.91e-28      Length:          321
Score:                  427.00        Matches:         99
Percent Similarity:     47.35%        Conservative:    53
Best Local Similarity:  30.84%        Mismatches:      87
Query Match:            31.28%        Indels:          82
DB:                     1             Gaps:            11

US-10-054-536-2 (1-747) x CL43_BOVIN (1-321)

Qy         19 CTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAG 78
              |||||||||   |||      ::::::      :::   |||::::::          |||
Db          2 LeuProLeuProLeuSerIleLeuLeuLeuLeuThrGlnSerGlnSerPheLeuGlyGlu 21

Qy         79 GAT-----------GCCCAAAAGACC-----------TGCCCTGCAGTGATTGCCTGT 114
              :::           ::::::|||||           |||      |||:::   |||
Db         22 GluMetAspValTyrSerGluLysThrLeuThrAspProCysThrLeuValVal---Cys 40

Qy        115 AGCTCTCCA------GGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAG 168
              :::   |||         :::   |||      |||:::||||||::::::|||   :::
Db         41 AlaProProAlaAspSerLeuArgGlyHisAspGlyArgAspGlyLysGluGlyProGln 60

Qy        169 GGAGAAAAGGGGGAACCAGGC-----------CAAGGGCTCAGAGGCTTACAGGGCCCC 216
              ||||||||||||:::||||||             |||      |||   :::||||||
Db         61 GlyGluLysGlyAspProGlyProProGlyMetProGlyProAlaGlyArgGluGlyPro 80

Qy        217 CCTGGAAAG---------TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGA 267
              |||:::          :::||||||||||||   |||||||||   |||   ||||||
Db         81 SerGlyArgGlnGlySerMetGlyProProGlyThrProGlyProLysGlyGluProGly 100

Qy        268 CCA---------------------------------------------------------- 270
              |||
Db        101 ProGluGlyGlyValGlyAlaProGlyMetProGlySerProGlyProAlaGlyLeuLys 120

Qy        270 ---------------------------------------------------------- 270

Db        121 GlyGluArgGlyThrProGlyProGlyGlyAlaIleGlyProGlnGlyProSerGlyAla 140

Qy        271 -----------------AAGGGCCAAAAAGGAGACCCTGGA---AAAAGTCCGGATGGT 309
                              ||||||   :::||||||||||||   |||         |||
Db        141 MetGlyProProGlyLeuLysGlyAspArgGlyAspProGlyGluLysGlyAlaArgGly 160

Qy        310 GATAGTAGCCTG------------------------------------GCTGCCTCAGAA 333
              ::::::|||:::                                    |||
Db        161 GluThrSerValLeuGluValAspThrLeuArgGlnArgMetArgAsnLeuGluGlyGlu 180

Qy        334 AGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGC 393
              :::   ||||||      :::   :::   :::|||   :::   |||      |||
Db        181 ValGlnArgLeuGlnAsnIleValThrGlnTyrArgLysAlaValLeuPheProAspGly 200

Qy        394 AAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTG 453
              :::   ||||||   |||   |||   |||   |||   :::   ::::::
Db        201 GlnAlaValGlyGluLysIlePheLysThrAlaGlyAlaValLysSerTyrSerAspAla 220

Qy        454 AAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAAT 513
              :::   ||||||   :::   :::      :::|||:::||||||::::::|||||||||
Db        221 GluGlnLeuCysArgGluAlaLysGlyGlnLeuAlaSerProArgSerSerAlaGluAsn 240

Qy        514 GGAGCCATTCAGAATCTCATCAAG---------GAGGAAGCCTTCCTGGGCATCACTGAT 564
                 |||:::      |||::::::          :::   |||:::|||   :::   |||
Db        241 GluAlaValThrGlnLeuValArgAlaLysAsnLysHisAlaTyrLeuSerMetAsnAsp 260

Qy        565 GAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGG 624
                    ||||||:::|||          ||||||      |||   |||:::||||||
Db        261 IleSerLysGluGlyLysPheThrTyrProThrGlyGlySerLeuAspTyrSerAsnTrp 280

Qy        625 AACGAGGGTGAACCCAACAATGCTGGTTCTGAT---------GAAGATTGTGTATTGCTA 675
                  ||||||||||||||||          |||          |||:::|||:::   :::
Db        281 AlaProGlyGluProAsnAsnArgAlaLysAspGluGlyProGluAsnCysLeuGluIle 300

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
                  :::|||   |||||||||:::   |||             |||   :::||||||
Db        301 TyrSerAspGlyAsnTrpAsnAspIleGluCysArgGluGluArgLeuValIleCysGlu 320

Qy        736 TTC 738
              |||
Db        321 Phe 321

RESULT 13
PSPA_RABIT
ID   PSPA_RABIT     STANDARD;      PRT;   247 AA.
AC   P12842;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE   (PSAP).
GN   SFTPA1 OR SFTPA OR SFTP1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88139348; PubMed=2830270;
RA   Boggaram V., Qing K., Mendelson C.R.;
RT   "The major apoprotein of rabbit pulmonary surfactant. Elucidation of
RT   primary sequence and cyclic AMP and developmental regulation.";
RL   J. Biol. Chem. 263:2939-2947(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=New Zealand white; TISSUE=Liver;
RX   MEDLINE=92312742; PubMed=1616051;
RA   Chen Q., Boggaram V., Mendelson C.R.;
RT   "Rabbit lung surfactant protein A gene: identification of a lung-
RT   specific DNase I hypersensitive site.";
RL   Am. J. Physiol. 262:L662-L671(1992).
CC   -!- FUNCTION: In presence of calcium ions, PSAP binds to surfactant
CC       phospholipids and contributes to lower the surface tension at the
CC       air-liquid interface in the alveoli of the mammalian lung and is
CC       essential for normal respiration.
CC   -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- MISCELLANEOUS: Pulmonary surfactant consists of 90% lipid and 10%
CC       protein. There are 4 surfactant-associated protein: 2 collagenous,
CC       carbohydrate-binding glycoproteins (SP-A and SP-D) and 2 small
CC       hydrophobic proteins (SP-B and SP-C).
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J03542; AAA31465.1; -.
DR   EMBL; L19387; AAA31468.1; -.
```

```
DR   PIR; A29931; LNRBPS.
DR   HSSP; P22897; 1EGG.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW   Signal; Lectin; Collagen; Repeat.
FT   SIGNAL       1    15       POTENTIAL.
FT   CHAIN       16   247       PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT                              A.
FT   DOMAIN      27    99       COLLAGEN-LIKE.
FT   DOMAIN     152   247       C-TYPE LECTIN (SHORT FORM).
FT   DISULFID   154   245       BY SIMILARITY.
FT   DISULFID   223   237       BY SIMILARITY.
FT   CARBOHYD   206   206       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   VARIANT     12    12       S -> P.
FT   CONFLICT    57    60       GPMG -> APWA (IN REF. 2).
SQ   SEQUENCE   247 AA;  26071 MW;  289634054CBC8CB4 CRC64;

Alignment Scores:
Pred. No.:              4.04e-27      Length:          247
Score:                  417.50        Matches:          91
Percent Similarity:     52.17%        Conservative:     41
Best Local Similarity:  35.97%        Mismatches:       94
Query Match:            30.59%        Indels:           27
DB:                     1             Gaps:              6

US-10-054-536-2 (1-747) x PSPA_RABIT (1-247)

Qy         28 CTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCCAA 87
              |||   |||:::::::::::::|||      |||:::|||         :::|||
Db          6 LeuAlaLeuThrLeuIleSerAlaProAlaSerAspThrCysAspThrLysAspValCys 25

Qy         88 AAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAA 147
                 |||   :::           :::||||||   :::|||   ||||||:::
Db         26 IleGlySerProGlyIle------ProGlyThrProGlySerHisGlyLeuProGlyArg 43

Qy        148 GATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAA---------------- 192
              ||||||||||||||   ||||||:::   |||   ||||||
Db         44 AspGlyArgAspGlyValLysGlyAspProGlyProProGlyProMetGlyProProGly 63

Qy        193 GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCT 252
              |||:::   ||||||   |||       |||   :::|||   ||||||   ||||||
Db         64 GlyMetProGlyLeuProGlyArgAspGlyLeuIleGlyAlaProGlyValProGlyGlu 83

Qy        253 TCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGAT 312
                 |||       |||       |||::::::|||   ||||||
Db         84 ArgGlyAspLysGlyGluProGlyGluArgGlyProProGly------------------ 97

Qy        313 AGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAG 372
                 |||   |||         :::   ||||||      :::      ::::::
Db         98 ------LeuProAlaTyrLeuAspGluGluLeuGlnAlaThrLeuHisGluLeuArgHis 115

Qy        373 TGGCTGACCTTCTCTCTGGGC----------------------AAACAAGTTGGGAAC 408
                       |||:::|||                          |||   ||||||
Db        116 HisAlaLeuGlnSerIleGlyValLeuSerLeuGlnGlySerMetLysAlaValGlyGlu 135

Qy        409 AAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTC 468
              |||   |||   |||||||||:::   :::   |||:::   ::::::   :::|||
Db        136 LysIlePheSerThrAsnGlyGlnSerValAsnPheAspAlaIleArgGluValCysAla 155

Qy        469 AAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAAT 528
              :::           :::|||   ||||||:::      ||||||   ||||||   :::
Db        156 ArgAlaGlyGlyArgIleAlaValProArgSerLeuGluGluAsnGluAlaIleAlaSer 175

Qy        529 CTCATCAAGGAGGAA---------GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGG 579
              ::::::||||||             |||:::||||||:::   :::      |||   |||
Db        176 IleValLysGluArgAsnThrTyrAlaTyrLeuGlyLeuAlaGluGlyProThrAlaGly 195

Qy        580 CAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCC 639
                 |||       |||   |||:::   :::   ||||||||||||      |||||||||
Db        196 AspPheTyrTyrLeuAspGlyAspProValAsnTyrThrAsnTrpTyrProGlyGluPro 215

Qy        640 AACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGAC 699
                       |||       |||   ||||||   :::      :::|||:::|||||||||
Db        216 ArgGlyGlnGlyArg---GluLysCysValGluMetTyrThrAspGlyLysTrpAsnAsp 234

Qy        700 GTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
                       |||            |||   :::|||||||||
Db        235 LysAsnCysLeuGlnTyrArgLeuValIleCysGluPhe 247

RESULT 14
PSPD_BOVIN
ID   PSPD_BOVIN     STANDARD;      PRT;   369 AA.
AC   P35246;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN   SFTPD OR SFTP4.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
RC   TISSUE=Lung;
RX   MEDLINE=93170856; PubMed=8436402;
RA   Lim B.L., Lu J., Reid K.B.M.;
RT   "Structural similarity between bovine conglutinin and bovine lung
RT   surfactant protein D and demonstration of liver as a site of
RT   synthesis of conglutinin.";
RL   Immunology 78:159-165(1993).
CC   -!- FUNCTION: Contributes to the lung's defense against inhaled
CC       microorganisms. Binds strongly maltose residues and to a lesser
CC       extent other alpha-glucosyl moieties. It could participate in the
CC       extracellular reorganization or turnover of pulmonary surfactant.
CC   -!- SUBUNIT: Oligomeric complex of 4 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- MISCELLANEOUS: Pulmonary surfactant consists of 90% lipid and 10%
CC       protein. There are 4 surfactant-associated protein: 2 collagenous,
CC       carbohydrate-binding glycoproteins (SP-A and SP-D) and 2 small
CC       hydrophobic proteins (SP-B and SP-C).
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75911; CAA53510.1; -.
DR   HSSP; P35247; 1B08.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW   Signal; Lectin; Collagen; Repeat; Coiled coil.
FT   SIGNAL       1    20       BY SIMILARITY.
FT   CHAIN       21   369       PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT                              D.
FT   DOMAIN      46   216       COLLAGEN-LIKE.
FT   DOMAIN     217   248       COILED COIL (POTENTIAL).
FT   DOMAIN     273   369       C-TYPE LECTIN (SHORT FORM).
```

```
FT   DISULFID    275    367       BY SIMILARITY.
FT   DISULFID    345    359       BY SIMILARITY.
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MOD_RES      78     78       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES      87     87       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES      96     96       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES      99     99       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     165    165       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     171    171       HYDROXYLATION (BY SIMILARITY).
SQ   SEQUENCE   369 AA;  37361 MW;  07D88B24E0AEB2E3 CRC64;

Alignment Scores:
Pred. No.:                6.47e-27      Length:          369
Score:                    415.50        Matches:         87
Percent Similarity:       52.23%        Conservative:    30
Best Local Similarity:    38.84%        Mismatches:      90
Query Match:              30.44%        Indels:          17
DB:                       1             Gaps:            4

US-10-054-536-2 (1-747) x PSPD_BOVIN (1-369)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::|||||||||   |||   ||||||      |||      |||   :::|||
Db        146 AlaProGlyIleGlnGlySerProGlyProAlaGlyLeuLysGlyGluArgGlyAlaPro 165

Qy        178 GGGGAACCAGGC---CAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCT 234
              ||||||||||||      |||      |||      ||||||   |||   :::||||||
Db        166 GlyGluProGlyAlaProGlyArgAlaGlyAlaProGlyProAlaGlyAlaIleGlyPro 185

Qy        235 CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA 294
              |||      |||      |||   ||||||   ||||||   :::|||   ||||||
Db        186 GlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGlyThrProGly 205

Qy        295 ---AAAAGTCCGGATGGTGATAGTAGCCTGGCT-------------------------- 324
                  :::        |||:::|||   ||||||
Db        206 GluArgGlyAlaLysGlyGluSerGlyLeuAlaGluValAsnAlaLeuArgGlnArgVal 225

Qy        325 ---------GCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGG 375
                      :::   :::   ||||||        ::::::   ||||||
Db        226 GlyIleLeuGluGlyGlnLeuGlnArgLeuGlnAsnAlaPheSerGlnTyrLysLysAla 245

Qy        376 CTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATA 435
              :::   |||      |||:::   ||||||   |||   |||   |||   |||
Db        246 MetLeuPheProAsnGlyArgSerValGlyGluLysIlePheLysThrValGlySerGlu 265

Qy        436 ATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCC 495
              ||||||:::      :::   :::|||   :::          :::   :::|||
Db        266 LysThrPheGlnAspAlaGlnGlnIleCysThrGlnAlaGlyGlyGlnLeuProSerPro 285

Qy        496 AGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAGGAGGAAGCC 546
              |||:::   |||||||||   |||:::      |||                :::   |||
Db        286 ArgSerGlyAlaGluAsnGluAlaLeuThrGlnLeuAlaThrAlaGlnAsnLysAlaAla 305

Qy        547 TTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGA 606
              ||||||   ::::::|||   :::   ||||||   |||:::      ||||||
Db        306 PheLeuSerMetSerAspThrArgLysGluGlyThrPheIleTyrProThrGlyGluPro 325

Qy        607 CTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGT 666
              |||   |||:::||||||          ||||||||||||   |||      |||:::|||
Db        326 LeuValTyrSerAsnTrpAlaProGlnGluProAsnAsnAspGlyGlySerGluAsnCys 345

Qy        667 GTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCC 726
              |||   :::      ||||||:::||||||||      |||      |||
Db        346 ValGluIlePheProAsnGlyLysTrpAsnAspLysValCysGlyGluGlnArgLeuVal 365

Qy        727 GTCTGTGAGTTC 738
              :::|||||||||
Db        366 IleCysGluPhe 369

RESULT 15
PSPA_CAVPO
ID   PSPA_CAVPO     STANDARD;      PRT;   247 AA.
AC   P50403;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE   (PSAP).
GN   SFTPA1 OR SFTPA OR SFTP1.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hartley; TISSUE=Lung;
RX   MEDLINE=98018900; PubMed=9357868;
RA   Yuan H.T., Gowan S., Kelly F.J., Bingle C.D.;
RT   "Cloning of guinea pig surfactant protein A defines a distinct
RT   cellular distribution pattern within the lung.";
RL   Am. J. Physiol. 273:L900-L906(1997).
CC   -!- FUNCTION: In presence of calcium ions, PSAP binds to surfactant
CC       phospholipids and contributes to lower the surface tension at the
CC       air-liquid interface in the alveoli of the mammalian lung and is
CC       essential for normal respiration.
CC   -!- SUBUNIT: Oligomeric complex of 6 set of homotrimers.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- MISCELLANEOUS: Pulmonary surfactant consists of 90% lipid and 10%
CC       protein. There are 4 surfactant-associated protein: 2 collagenous,
CC       carbohydrate-binding glycoproteins (SP-A and SP-D) and 2 small
CC       hydrophobic proteins (SP-B and SP-C).
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U40869; AAB82952.1; -.
DR   HSSP; P22897; 1EGG.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW   Signal; Lectin; Collagen; Repeat.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    247       PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT                                A.
FT   DOMAIN       27     99       COLLAGEN-LIKE.
FT   DOMAIN      152    245       C-TYPE LECTIN (SHORT FORM).
FT   DISULFID    154    245       BY SIMILARITY.
FT   DISULFID    223    237       BY SIMILARITY.
FT   CARBOHYD     20     20       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    206    206       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   247 AA;  26104 MW;  D1BC86270EEFC932 CRC64;

Alignment Scores:
Pred. No.:                4.04e-26      Length:          247
Score:                    405.50        Matches:         98
Percent Similarity:       52.71%        Conservative:    38
Best Local Similarity:    37.98%        Mismatches:      89
Query Match:              29.71%        Indels:          33
DB:                       1             Gaps:            10

US-10-054-536-2 (1-747) x PSPA_CAVPO (1-247)
```

```
Qy         16 TCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGT  75
              :::|||   |||:::|||::::::::::                :::   |||      |||
Db          6 AlaLeuIleLeuIleLeuValThrValSerGlyIleMetCysAsnArgThrAspPheCys  25

Qy         76 GAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGC 135
                 :::           |||   :::           :::||||||   :::|||
Db         26 ValGlySer----------ProGlyIle------ProGlyThrProGlySerHisGly  39

Qy        136 TTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGG 195
                ||||||:::|||||||||||||||   ||||||:::   |||   ||||||
Db         40 LeuProGlyArgAspGlyArgAspGlyValLysGlyAspProGlyProProGlyPro---  58

Qy        196 CTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCT 255
                           ||||||||||||   :::      ||||||   ||||||   :::
Db         59 ------------MetGlyProProGlyValMet------ProGlyPheProGlyCysAsn  72

Qy        256 ------------------GGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAA 297
                          |||:::||||||   :::|||   |||||||||||||||:::
Db         73 GlyMetAsnGlyIleProGlyAlaProGlyGluArgGlyAspLysGlyAspProGlyGlu  92

Qy        298 AGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATG 357
                  |||          |||   ||||||      :::   :::||||||
Db         93 Arg------GlyProProGlyLeuProAlaSerLeuAspGluGluIleGlnThrIlePhe 110

Qy        358 GCACGTATCAAAAAG---------TGGCTGACCTTCTCTCTGGGCAAACAA--------- 399
                 :::|||               ||||||      |||      |||
Db        111 HisAsnLeuLysHisLysIleLeuGlnLeuThrGlyValLeuSerLeuGlnGlySerMet 130

Qy        400 ------GTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTG 453
                    ||||||:::|||   |||   |||||||||:::   :::   |||      :::
Db        131 LeuAlaValGlyAspLysValPheAlaThrAsnGlyGlnSerValAspPheAsnAlaIle 150

Qy        454 AAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAAT 513
                 |||      |||   :::             ||||||   |||||||||:::   ||||||
Db        151 LysGluThrCysAlaArgAlaGlyGlyAspValAlaAlaProArgAsnSerGluGluAsn 170

Qy        514 GGAGCCATTCAGAATCTCATCAAGGAG---------GAAGCCTTCCTGGGCATCACTGAT 564
                 ||||||   :::::::::|||:::           ::::::|||||:::|||:::
Db        171 ThrAlaIleSerSerIleValLysLysTyrAsnIleTyrSerTyrLeuGlyLeuThrGlu 190

Qy        565 GAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGG 624
                  |||   |||   |||      |||   |||:::   |||   ||||||||||||
Db        191 GlyHisThrProGlyAspPheHisTyrLeuAspGlySerProLeuAsnTyrThrAsnTrp 210

Qy        625 AACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAAT 684
                 ||||||||         |||      |||   |||      :::      :::
Db        211 TyrProGlyGluProArgGlyArgGlyLys---GluLysCysAlaGluMetTyrLeuAsp 229

Qy        685 GGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
                 |||   |||||||||      |||      |||   |||   :::|||||||||
Db        230 GlyThrTrpAsnAspLysAsnCysLeuGlnSerArgLeuThrIleCysGluPhe 247


Search completed: June 18, 2004, 21:00:53
Job time : 20 secs
```

Blank

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 21, 2004, 16:39:39 ; Search time 3234 Seconds
(without alignments)
10011.517 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 747
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	747	100.0	747	6	E27637	E27637 Recombinant
	2	747	100.0	900	6	AR182149	AR182149 Sequence
	3	747	100.0	1632	9	HOSA16576	Y16576 Homo sapien
	4	747	100.0	1638	9	HOSA16577	Y16577 Homo sapien
	5	747	100.0	1638	9	HOSA16580	Y16580 Homo sapien
	6	747	100.0	1638	9	HOSA16581	Y16581 Homo sapien
	7	747	100.0	3605	6	E27636	E27636 Recombinant
	8	747	100.0	3605	6	AX411061	AX411061 Sequence
	9	747	100.0	3605	9	HSMBPC	X15422 Human mRNA
	10	745.4	99.8	1632	9	HOSA16578	Y16578 Homo sapien
	11	745.4	99.8	1638	9	HOSA16579	Y16579 Homo sapien
	12	745.4	99.8	1638	9	HOSA16582	Y16582 Homo sapien
	13	743.8	99.6	776	9	AF360991	AF360991 Homo sapi
	14	635.8	85.1	1188	6	I09220	I09220 Sequence 1
	15	587.8	78.7	805	9	MACMBPC	L43911 Macaca mula
	16	444.6	59.5	1409	4	D73408	D73408 Bos taurus
	17	406.6	54.4	723	4	AF164576	AF164576 Sus scrof
	18	395	52.9	1010	6	E37364	E37364 Swine serum
	19	381.4	51.1	799	9	MACMBPA	L43912 Macaca mula
	20	374.8	50.2	1116	10	S42294	S42294 mannose-bin
	21	374	50.1	1254	9	HSMBP1D	X15957 H.sapiens M
	22	374	50.1	3336	9	HSMBPCA3	AF080510 Homo sapi
	23	373.2	50.0	1069	10	MUSRRFA	D11440 Mus musculu
	24	373.2	50.0	1098	10	BC010760	BC010760 Mus muscu
c	25	372.4	49.9	102532	9	AL731550	AL731550 Human DNA
c	26	370.8	49.6	302859	2	AC044785	AC044785 Homo sapi
	27	366.2	49.0	874	10	RNMBPR	X05023 Rat mRNA fo
	28	364.6	48.8	1037	6	AX402034	AX402034 Sequence
	29	364.6	48.8	1037	10	RATMABPC	M14103 Rat mannose
	30	287.6	38.5	1068	6	AX827373	AX827373 Sequence
	31	287.6	38.5	1068	10	AF080507	AF080507 Rattus sp
	32	286	38.3	866	10	MUSRRFB	D11441 Mus musculu
	33	286	38.3	894	10	BC021762	BC021762 Mus muscu
	34	284.4	38.1	943	6	E63781	E63781 Serum manna
	35	284.4	38.1	943	10	S42292	S42292 mannose-bin
	36	280.6	37.6	1423	9	AF019382	AF019382 Homo sapi
c	37	206	27.6	141678	10	AC102722	AC102722 Mus muscu
	38	204.4	27.4	1211	6	AR182169	AR182169 Sequence
	39	204.4	27.4	1211	10	MMMBII6	U09016 Mus musculu
c	40	204.4	27.4	223274	2	AC112443	AC112443 Rattus no
	41	204.4	27.4	270783	2	AC105489	AC105489 Rattus no
c	42	193.8	25.9	190092	2	AC120502	AC120502 Lemur cat
	43	192	25.7	134411	9	AL512662	AL512662 Human DNA
	44	188	25.2	1110	9	HSMBP1A	X15954 H.sapiens M
	45	188	25.2	1802	9	HSMBPCA1	AF080508 Homo sapi

ALIGNMENTS

RESULT 1
E27637
LOCUS E27637 747 bp DNA linear PAT 18-JUN-2001
DEFINITION Recombinant human mannan binding protein and process for producing the same.
ACCESSION E27637
VERSION E27637.1 GI:13018239
KEYWORDS JP 1999206378-A/2.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 747)
AUTHORS Nobutaka,W.
TITLE Recombinant human mannan binding protein and process for producing the same

```
  JOURNAL    Patent: JP 1999206378-A 2 03-AUG-1999;
             FUSO YAKUHIN KOGYO KK
COMMENT      OS   Unidentified
             PN   JP 1999206378-A/2
             PD   03-AUG-1999
             PF   23-JAN-1998 JP 1998011864
             PR
             PI   NOBUTAKA WAKAMIYA
             PC   C12N15/09,C07K14/47,C12P21/02//A61K38/00,(C12N15/09,C12R1:91),
             PC   (C12P21/02,C12R1:91),C12N15/00,A61K37/02,(C12N15/00,C12R1:91)
             CC   Strandedness: Double;
             CC   Topology: Linear;
             FH   Key             Location/Qualifiers
             FT   source          1. .747
             FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 747;  DB 6;  Length 747;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747


RESULT 2
AR182149
LOCUS       AR182149                900 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6337193.
ACCESSION   AR182149
VERSION     AR182149.1  GI:20225065
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   Tully,R.E., Caltagirone,G.Thomas., Moyer,S.S. and Ronning,M.T.
  TITLE     Expression of manose-binding protein in methylotrophic yeast
  JOURNAL   Patent: US 6337193-A 1 08-JAN-2002;
FEATURES             Location/Qualifiers
     source          1. .900
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 747;  DB 6;  Length 900;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725
```

```
Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812


RESULT 3
HOSA16576
LOCUS       HOSA16576               1632 bp    DNA     linear   PRI 17-SEP-1999
DEFINITION  Homo sapiens gene encoding mannan/mannose-binding protein, variant
            LYQA.
ACCESSION   Y16576
VERSION     Y16576.1  GI:5911789
KEYWORDS    mannose-binding lectin; mbl gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madsen,H.O., Satz,M.L., Hogh,B., Svejgaard,A. and Garred,P.
  TITLE     Different molecular events result in low protein levels of
            mannan-binding lectin in populations from southeast Africa and
            South America
  JOURNAL   J. Immunol. 161 (6), 3169-3175 (1998)
  MEDLINE   98414317
   PUBMED   9743385
REFERENCE   2  (bases 1 to 1632)
  AUTHORS   Madsen,H.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) H.O. Madsen, Department of Clinical
            Immunology, Section 7631, National University Hospital, Tagensvej
            20, DK-2200 Copenhagen, DENMARK
COMMENT     Related sequences  X15954, X15955, X15956, X15422.
FEATURES             Location/Qualifiers
     source          1. .1632
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /note="MBL haplotype LYQA"
     variation       273
                     /note="polymorphism in different MBL haplotypes"
     variation       396
                     /note="polymorphism in different MBL haplotypes"
     variation       474
                     /note="polymorphism in different MBL haplotypes"
     variation       487
                     /note="polymorphism in different MBL haplotypes"
     variation       596
                     /note="polymorphism in different MBL haplotypes"
     gene            739. .1632
                     /gene="MBL"
     CAAT_signal     739. .744
                     /gene="MBL"
     variation       747
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     TATA_signal     780. .785
                     /gene="MBL"
     mRNA            817. .>1632
                     /gene="MBL"
                     /product="mannose-binding lectin"
     variation       820
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     CDS             886. .1632
                     /gene="MBL"
                     /codon_start=1
                     /product="mannose-binding lectin"
                     /protein_id="CAB56044.1"
                     /db_xref="GI:5911790"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDGTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     886. .945
                     /gene="MBL"
     variation       1039
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1046
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1055
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
ORIGIN

  Query Match             100.0%;  Score 747;  DB 9;  Length 1632;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        886 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 945

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        946 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 1005

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1006 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 1065

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1066 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 1125

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1126 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 1185

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1186 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 1245

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1246 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 1305

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1306 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 1365

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 1425

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 1485

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1486 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 1545

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1546 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 1605

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              ||||||||||||||||||||||||||||
Db       1606 CTGGCCGTCTGTGAGTTCCCTATCTGA 1632


RESULT 4
HOSA16577
LOCUS       HOSA16577               1638 bp    DNA     linear   PRI 17-SEP-1999
DEFINITION  Homo sapiens gene encoding mannan/mannose-binding protein, variant
            LYPA.
ACCESSION   Y16577
VERSION     Y16577.1  GI:5911791
KEYWORDS    mannose-binding lectin; mbl gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madsen,H.O., Satz,M.L., Hogh,B., Svejgaard,A. and Garred,P.
  TITLE     Different molecular events result in low protein levels of
            mannan-binding lectin in populations from southeast Africa and
            South America
  JOURNAL   J. Immunol. 161 (6), 3169-3175 (1998)
  MEDLINE   98414317
   PUBMED   9743385
REFERENCE   2  (bases 1 to 1638)
  AUTHORS   Madsen,H.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) H.O. Madsen, Department of Clinical
            Immunology, Section 7631, National University Hospital, Tagensvej
            20, DK-2200 Copenhagen, DENMARK
COMMENT     Related sequences  X15954, X15955, X15956, X15422.
FEATURES             Location/Qualifiers
     source          1. .1638
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /note="MBL haplotype LYPA"
     variation       273
                     /note="polymorphism in different MBL haplotypes"
     variation       396
                     /note="polymorphism in different MBL haplotypes"
     variation       474
                     /note="polymorphism in different MBL haplotypes"
     variation       487
                     /note="polymorphism in different MBL haplotypes"
     variation       495. .500
                     /note="polymorphism in different MBL haplotypes"
     gene            745. .1638
                     /gene="MBL"
     CAAT_signal     745. .750
                     /gene="MBL"
     variation       753
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     TATA_signal     786. .791
                     /gene="MBL"
     mRNA            823. .>1638
                     /gene="MBL"
                     /product="mannose-binding lectin"
     variation       826
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     CDS             892. .1638
                     /gene="MBL"
                     /codon_start=1
                     /product="mannose-binding lectin"
                     /protein_id="CAB56120.1"
                     /db_xref="GI:5911792"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDGTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     892. .951
                     /gene="MBL"
     variation       1045
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1052
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1061
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
ORIGIN

  Query Match             100.0%;  Score 747;  DB 9;  Length 1638;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 951

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 1011

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1012 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 1071

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 1131

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 1191

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1192 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 1251

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1252 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 1311

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1312 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 1371

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1372 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 1431

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1432 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 1491

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1492 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 1551

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1552 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 1611

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
```

```
                    ||||||||||||||||||||||||||
Db       1612 CTGGCCGTCTGTGAGTTCCCTATCTGA 1638


RESULT 5
HOSA16580
LOCUS       HOSA16580               1638 bp    DNA     linear   PRI 17-SEP-1999
DEFINITION  Homo sapiens gene encoding mannan/mannose-binding protein, variant
            LXPA.
ACCESSION   Y16580
VERSION     Y16580.1  GI:5911797
KEYWORDS    mannose-binding lectin; mbl gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madsen,H.O., Satz,M.L., Hogh,B., Svejgaard,A. and Garred,P.
  TITLE     Different molecular events result in low protein levels of
            mannan-binding lectin in populations from southeast Africa and
            South America
  JOURNAL   J. Immunol. 161 (6), 3169-3175 (1998)
  MEDLINE   98414317
   PUBMED   9743385
REFERENCE   2  (bases 1 to 1638)
  AUTHORS   Madsen,H.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) H.O. Madsen, Department of Clinical
            Immunology, Section 7631, National University Hospital, Tagensvej
            20, DK-2200 Copenhagen, DENMARK
COMMENT     Related sequences  X15954, X15955, X15956, X15422.
FEATURES             Location/Qualifiers
     source          1. .1638
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /note="MBL haplotype LXPA"
variation            273
                     /note="polymorphism in different MBL haplotypes"
variation            396
                     /note="polymorphism in different MBL haplotypes"
variation            474
                     /note="polymorphism in different MBL haplotypes"
variation            487
                     /note="polymorphism in different MBL haplotypes"
variation            495. .500
                     /note="polymorphism in different MBL haplotypes"
variation            602
                     /note="polymorphism in different MBL haplotypes"
gene                 745. .1638
                     /gene="MBL"
CAAT_signal          745. .750
                     /gene="MBL"
variation            753
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
TATA_signal          786. .791
                     /gene="MBL"
mRNA                 823. .>1638
                     /gene="MBL"
                     /product="mannose-binding lectin"
variation            826
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
CDS                  892. .1638
                     /gene="MBL"
                     /codon_start=1
                     /product="mannose-binding lectin"
                     /protein_id="CAB56122.1"
                     /db_xref="GI:5911798"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDGTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     892. .951
                     /gene="MBL"
     variation       1045
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1052
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1061
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
ORIGIN

  Query Match           100.0%;  Score 747;  DB 9;  Length 1638;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 951

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 1011

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1012 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 1071

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 1131

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 1191

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1192 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 1251

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1252 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 1311

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1312 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 1371

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1372 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 1431

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1432 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 1491

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1492 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 1551

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1552 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 1611

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
```

```
Db       1612 CTGGCCGTCTGTGAGTTCCCTATCTGA 1638


RESULT 6
HOSA16581
LOCUS       HOSA16581               1638 bp    DNA     linear   PRI 17-SEP-1999
DEFINITION  Homo sapiens gene encoding mannan/mannose-binding protein, variant
            HYPA.
ACCESSION   Y16581
VERSION     Y16581.1  GI:5911806
KEYWORDS    mannose-binding lectin; mbl gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madsen,H.O., Satz,M.L., Hogh,B., Svejgaard,A. and Garred,P.
  TITLE     Different molecular events result in low protein levels of
            mannan-binding lectin in populations from southeast Africa and
            South America
  JOURNAL   J. Immunol. 161 (6), 3169-3175 (1998)
  MEDLINE   98414317
   PUBMED   9743385
REFERENCE   2  (bases 1 to 1638)
  AUTHORS   Madsen,H.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) H.O. Madsen, Department of Clinical
            Immunology, Section 7631, National University Hospital, Tagensvej
            20, DK-2200 Copenhagen, DENMARK
COMMENT     Related sequences  X15954, X15955, X15956, X15422.
FEATURES             Location/Qualifiers
     source          1. .1638
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /note="MBL haplotype HYPA"
     variation       273
                     /note="polymorphism in different MBL haplotypes"
     variation       396
                     /note="polymorphism in different MBL haplotypes"
     variation       474
                     /note="polymorphism in different MBL haplotypes"
     variation       487
                     /note="polymorphism in different MBL haplotypes"
     variation       495. .500
                     /note="polymorphism in different MBL haplotypes"
     variation       602
                     /note="polymorphism in different MBL haplotypes"
     gene            745. .1638
                     /gene="MBL"
     CAAT_signal     745. .750
                     /gene="MBL"
     variation       753
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     TATA_signal     786. .791
                     /gene="MBL"
     mRNA            823. .>1638
                     /gene="MBL"
                     /product="mannose-binding lectin"
     variation       826
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     CDS             892. .1638
                     /gene="MBL"
                     /codon_start=1
                     /product="mannose-binding lectin"
                     /protein_id="CAB56123.1"
                     /db_xref="GI:5911807"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDGTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     892. .951
                     /gene="MBL"
     variation       1045
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1052
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1061
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
ORIGIN

  Query Match             100.0%;  Score 747;  DB 9;  Length 1638;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 951

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 1011

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1012 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 1071

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 1131

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 1191

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1192 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 1251

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1252 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 1311

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1312 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 1371

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1372 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 1431

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1432 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 1491

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1492 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 1551

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1552 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 1611

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db       1612 CTGGCCGTCTGTGAGTTCCCTATCTGA 1638
```

```
RESULT 7
E27636
LOCUS       E27636                  3605 bp    DNA     linear   PAT 18-JUN-2001
DEFINITION  Recombinant human mannan binding protein and process for producing
            the same.
ACCESSION   E27636
VERSION     E27636.1  GI:13018238
KEYWORDS    JP 1999206378-A/1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3605)
  AUTHORS   Nobutaka,W.
  TITLE     Recombinant human mannan binding protein and process for producing
            the same
  JOURNAL   Patent: JP 1999206378-A 1 03-AUG-1999;
            FUSO YAKUHIN KOGYO KK
COMMENT     OS   Homo sapiens
            PN   JP 1999206378-A/1
            PD   03-AUG-1999
            PF   23-JAN-1998 JP 1998011864
            PR
            PI   NOBUTAKA WAKAMIYA
            PC   C12N15/09,C07K14/47,C12P21/02//A61K38/00,(C12N15/09,C12R1:91),
            PC   (C12P21/02,C12R1:91),C12N15/00,A61K37/02,(C12N15/00,C12R1:91)
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   sig peptide     66. .125
            FT   CDS             66. .812.
FEATURES             Location/Qualifiers
     source          1. .3605
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           100.0%;  Score 747;  DB 6;  Length 3605;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812


RESULT 8
AX411061
LOCUS       AX411061                3605 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 3708 from Patent WO0229103.
ACCESSION   AX411061
VERSION     AX411061.1  GI:21443766
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 3708 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .3605
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. X15422"
ORIGIN

  Query Match           100.0%;  Score 747;  DB 6;  Length 3605;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-209;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
```

```
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812
```

```
RESULT 9
HSMBPC
LOCUS       HSMBPC                  3605 bp    mRNA    linear   PRI 31-MAR-1995
DEFINITION  Human mRNA for mannose-binding protein C.
ACCESSION   X15422
VERSION     X15422.1  GI:34486
KEYWORDS    mannose-binding protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3605)
  AUTHORS   Sastry,K., Herman,G.A., Day,L., Deignan,E., Bruns,G., Morton,C.C.
            and Ezekowitz,R.A.
  TITLE     The human mannose-binding protein gene. Exon structure reveals its
            evolutionary relationship to a human pulmonary surfactant gene and
            localization to chromosome 10
  JOURNAL   J. Exp. Med. 170 (4), 1175-1189 (1989)
  MEDLINE   90010778
   PUBMED   2477486
REFERENCE   2  (bases 1 to 3605)
  AUTHORS   Ezekowitz,R.A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-1989) Ezekowitz R.A.B., The Children's Hospital,
            Enders Building 7th Floor, 300 Longwood Avenue, Boston MA 02115, U
            S A
COMMENT     X15422 revises MBP cDNA seq published by:
             Ezekowitz et. al. J. Exp. Med. 167:1034-1046(1988). Data kindly
            reviewed (22-FEB-1990) by Ezekowitz R.A.B.
FEATURES             Location/Qualifiers
     source          1. .3605
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="10q11.2-q21."
                     /clone="48-11"
                     /cell_line="Hep62"
                     /cell_type="heatocytes"
                     /tissue_type="liver"
                     /clone_lib="lambda gt10"
     CDS             66. .812
                     /codon_start=1
                     /product="precursor protein"
                     /protein_id="CAA33462.1"
                     /db_xref="GI:34487"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDGTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     66. .125
                     /note="signal peptide (AA -20 to -1)"
     mat_peptide     126. .809
                     /product="mat. mannose-binding protein C (AA 1-228)"
ORIGIN
```

Query Match 100.0%; Score 747; DB 9; Length 3605;
Best Local Similarity 100.0%; Pred. No. 2.8e-209;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 125

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 185

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 245

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 305

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 365

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 425

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 485

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 545

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 605

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 665

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 725

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 785

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        786 CTGGCCGTCTGTGAGTTCCCTATCTGA 812
```

```
RESULT 10
HOSA16578
LOCUS       HOSA16578               1632 bp    DNA     linear   PRI 17-SEP-1999
DEFINITION  Homo sapiens gene encoding mannan/mannose-binding protein, variant
            LYQC.
ACCESSION   Y16578
VERSION     Y16578.1  GI:5911793
KEYWORDS    mannose-binding lectin; mbl gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madsen,H.O., Satz,M.L., Hogh,B., Svejgaard,A. and Garred,P.
  TITLE     Different molecular events result in low protein levels of
            mannan-binding lectin in populations from southeast Africa and
            South America
  JOURNAL   J. Immunol. 161 (6), 3169-3175 (1998)
  MEDLINE   98414317
   PUBMED   9743385
REFERENCE   2  (bases 1 to 1632)
  AUTHORS   Madsen,H.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) H.O. Madsen, Department of Clinical
            Immunology, Section 7631, National University Hospital, Tagensvej
            20, DK-2200 Copenhagen, DENMARK
COMMENT     Related sequences  X15954, X15955, X15956, X15422.
FEATURES             Location/Qualifiers
     source          1. .1632
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /note="MBL haplotype LYQC"
     variation       273
                     /note="polymorphism in different MBL haplotypes"
     variation       396
                     /note="polymorphism in different MBL haplotypes"
     variation       474
                     /note="polymorphism in different MBL haplotypes"
     variation       487
                     /note="polymorphism in different MBL haplotypes"
     variation       596
                     /note="polymorphism in different MBL haplotypes"
     gene            739. .1632
                     /gene="MBL"
     CAAT_signal     739. .744
                     /gene="MBL"
     variation       747
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     TATA_signal     780. .785
                     /gene="MBL"
     mRNA            817. .>1632
                     /gene="MBL"
                     /product="mannose-binding lectin"
     variation       820
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     CDS             886. .1632
                     /gene="MBL"
                     /codon_start=1
                     /product="mannose-binding lectin"
                     /protein_id="CAB56045.1"
                     /db_xref="GI:5911794"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDGTKEEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     886. .945
                     /gene="MBL"
     variation       1039
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1046
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1055
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
ORIGIN

  Query Match             99.8%;  Score 745.4;  DB 9;  Length 1632;
  Best Local Similarity   99.9%;  Pred. No. 7.5e-209;
  Matches  746;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        886 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 945

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        946 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 1005

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1006 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGAAGAAAAGGGG 1065

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1066 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 1125

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1126 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 1185

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1186 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 1245

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1246 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 1305

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1306 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 1365

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 1425

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 1485

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1486 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 1545

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1546 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 1605

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db       1606 CTGGCCGTCTGTGAGTTCCCTATCTGA 1632


RESULT 11
HOSA16579
```

```
LOCUS       HOSA16579               1638 bp    DNA     linear   PRI 17-SEP-1999
DEFINITION  Homo sapiens gene encoding mannan/mannose-binding protein, variant
            LYPB.
ACCESSION   Y16579
VERSION     Y16579.1  GI:5911795
KEYWORDS    mannose-binding lectin; mbl gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madsen,H.O., Satz,M.L., Hogh,B., Svejgaard,A. and Garred,P.
  TITLE     Different molecular events result in low protein levels of
            mannan-binding lectin in populations from southeast Africa and
            South America
  JOURNAL   J. Immunol. 161 (6), 3169-3175 (1998)
  MEDLINE   98414317
   PUBMED   9743385
REFERENCE   2  (bases 1 to 1638)
  AUTHORS   Madsen,H.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) H.O. Madsen, Department of Clinical
            Immunology, Section 7631, National University Hospital, Tagensvej
            20, DK-2200 Copenhagen, DENMARK
COMMENT     Related sequences  X15954, X15955, X15956, X15422.
FEATURES             Location/Qualifiers
     source          1. .1638
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /note="MBL haplotype LYPB"
     variation       273
                     /note="polymorphism in different MBL haplotypes"
     variation       396
                     /note="polymorphism in different MBL haplotypes"
     variation       474
                     /note="polymorphism in different MBL haplotypes"
     variation       487
                     /note="polymorphism in different MBL haplotypes"
     variation       495. .500
                     /note="polymorphism in different MBL haplotypes"
     gene            745. .1638
                     /gene="MBL"
     CAAT_signal     745. .750
                     /gene="MBL"
     variation       753
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     TATA_signal     786. .791
                     /gene="MBL"
     mRNA            823. .>1638
                     /gene="MBL"
                     /product="mannose-binding lectin"
     variation       826
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     CDS             892. .1638
                     /gene="MBL"
                     /codon_start=1
                     /product="mannose-binding lectin"
                     /protein_id="CAB56121.1"
                     /db_xref="GI:5911796"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDDTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     892. .951
                     /gene="MBL"
     variation       1045
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1052
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1061
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
ORIGIN
```

Query Match 99.8%; Score 745.4; DB 9; Length 1638;
Best Local Similarity 99.9%; Pred. No. 7.5e-209;
Matches 746; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 951

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 1011

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1012 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGACACCAAGGGAGAAAAGGGG 1071

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 1131

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 1191

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1192 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 1251

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1252 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 1311

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1312 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 1371

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1372 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 1431

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1432 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 1491

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1492 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 1551

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1552 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 1611

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db       1612 CTGGCCGTCTGTGAGTTCCCTATCTGA 1638
```

```
RESULT 12
HOSA16582
LOCUS       HOSA16582               1638 bp    DNA     linear   PRI 17-SEP-1999
DEFINITION  Homo sapiens gene encoding mannan/mannose-binding protein, variant
            HYPD.
```

```
ACCESSION   Y16582
VERSION     Y16582.1  GI:5911808
KEYWORDS    mannose-binding lectin; mbl gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madsen,H.O., Satz,M.L., Hogh,B., Svejgaard,A. and Garred,P.
  TITLE     Different molecular events result in low protein levels of
            mannan-binding lectin in populations from southeast Africa and
            South America
  JOURNAL   J. Immunol. 161 (6), 3169-3175 (1998)
  MEDLINE   98414317
   PUBMED   9743385
REFERENCE   2  (bases 1 to 1638)
  AUTHORS   Madsen,H.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) H.O. Madsen, Department of Clinical
            Immunology, Section 7631, National University Hospital, Tagensvej
            20, DK-2200 Copenhagen, DENMARK
COMMENT     Related sequences  X15954, X15955, X15956, X15422.
FEATURES             Location/Qualifiers
     source          1. .1638
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /note="MBL haplotype HYPD"
     variation       273
                     /note="polymorphism in different MBL haplotypes"
     variation       396
                     /note="polymorphism in different MBL haplotypes"
     variation       474
                     /note="polymorphism in different MBL haplotypes"
     variation       487
                     /note="polymorphism in different MBL haplotypes"
     variation       495. .500
                     /note="polymorphism in different MBL haplotypes"
     variation       602
                     /note="polymorphism in different MBL haplotypes"
     gene            745. .1638
                     /gene="MBL"
     CAAT_signal     745. .750
                     /gene="MBL"
     variation       753
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     TATA_signal     786. .791
                     /gene="MBL"
     mRNA            823. .>1638
                     /gene="MBL"
                     /product="mannose-binding lectin"
     variation       826
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     CDS             892. .1638
                     /gene="MBL"
                     /codon_start=1
                     /product="mannose-binding lectin"
                     /protein_id="CAB56124.1"
                     /db_xref="GI:5911809"
                     /db_xref="GOA:P11226"
                     /db_xref="SWISS-PROT:P11226"
                     /translation="MSLFPSLPLLLLSMVAASYSETVTCEDAQKTCPAVIACSSPGIN
                     GFPGKDGCDGTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     892. .951
                     /gene="MBL"
     variation       1045
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1052
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
     variation       1061
                     /gene="MBL"
                     /note="polymorphism in different MBL haplotypes"
ORIGIN
```

Query Match 99.8%; Score 745.4; DB 9; Length 1638;
Best Local Similarity 99.9%; Pred. No. 7.5e-209;
Matches 746; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 951

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 1011

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1012 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGTGTGATGGCACCAAGGGAGAAAAGGGG 1071

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 1131

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 1191

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1192 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 1251

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1252 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 1311

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1312 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 1371

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1372 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 1431

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1432 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 1491

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1492 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 1551

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1552 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 1611

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db       1612 CTGGCCGTCTGTGAGTTCCCTATCTGA 1638
```

```
RESULT 13
AF360991
LOCUS       AF360991                776 bp    mRNA    linear   PRI 14-MAY-2001
DEFINITION  Homo sapiens mannan-binding lectin MBL precursor (MBL) mRNA,
            complete cds.
ACCESSION   AF360991
```

```
VERSION     AF360991.1  GI:14030459
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   Chen,Z., Zhu,X. and Xie,P.
  TITLE     Cloning and sequencing of mannan-binding lectin cDNA of Chinese
  JOURNAL   Mianyixue Zazhi 15, 83-86 (1999)
REFERENCE   2  (bases 1 to 776)
  AUTHORS   Chen,Z., Zhu,X. and Xie,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2001) Immunology, First Military Medical
            University, Tonghe, Guangzhou, Guangdong 510515, P.R. of China
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q11.2-q21"
                     /cell_type="hepatocyte"
                     /tissue_type="liver"
     gene            1. .776
                     /gene="MBL"
     CDS             17. .763
                     /gene="MBL"
                     /note="activates complement system; serum protein"
                     /codon_start=1
                     /product="mannan-binding lectin MBL precursor"
                     /protein_id="AAK52907.1"
                     /db_xref="GI:14030460"
                     /translation="MSLFPSLPLLLLSMVAASYSETVACEDAQKTCPAVIACSSPGIN
                     GFPGKDGRDGTKGEKGEPGQGLRGLQGPPGKLGPPGNPGPSGSPGPKGQKGDPGKSPD
                     GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQA
                     SVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGEPNNAGS
                     DEDCVLLLKNGQWNDVPCSTSHLAVCEFPI"
     sig_peptide     17. .76
                     /gene="MBL"
     mat_peptide     77. .760
                     /gene="MBL"
                     /product="mannan-binding lectin MBL"
     variation       86
                     /gene="MBL"
                     /note="compared to Homo sapiens mannan-binding lectin
                     encoded by GenBank Accession Number X15422; Ala to Thr
                     substitution"
                     /replace="A"
     variation       571
                     /gene="MBL"
                     /note="compared to Homo sapiens mannan-binding lectin
                     encoded by GenBank Accession Number X15422; does not alter
                     sequence of predicted protein"
                     /replace="C"
ORIGIN
```

Query Match 99.6%; Score 743.8; DB 9; Length 776;
Best Local Similarity 99.7%; Pred. No. 2e-208;
Matches 745; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 76

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 GAAACTGTGGCCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 136

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 196

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCC
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCC 256

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 316

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 376

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 436

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 496

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 556

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        557 GAAGCCTTCCTGGGTATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 616

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 676

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 736

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              ||||||||||||||||||||||||||
Db        737 CTGGCCGTCTGTGAGTTCCCTATCTGA 763
```

```
RESULT 14
I09220
LOCUS       I09220                  1188 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 1 from Patent WO 8901519.
ACCESSION   I09220
VERSION     I09220.1  GI:588069
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1188)
  AUTHORS   Ezekowitz,R.A.B.
  TITLE     HUMAN MANNOSE BINDING PROTEIN
  JOURNAL   Patent: WO 8901519-A 1 23-FEB-1989;
FEATURES             Location/Qualifiers
     source          1. .1188
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Query Match 85.1%; Score 635.8; DB 6; Length 1188;
Best Local Similarity 95.7%; Pred. No. 1.8e-176;
Matches 707; Conservative 0; Mismatches 27; Indels 5; Gaps 5;

```
Qy         11 TTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGA 70
              || |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db         75 TTTCATCACTCCCTCT-CTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGA 133

Qy         71 CCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCA 130
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 CCTGTGAGGGTGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCA 193
```

```
Qy        131 ACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCC 190
              ||||||||||||||||||||||||||||||||||||||||||    ||| |||| | |||| |
Db        194 ACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGTAGAAAAGGGGGAACAGGTC 253

Qy        191 AAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGC 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 AAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGC 313

Qy        251 CTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 CTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTG 373

Qy        311 ATAGTAG-CCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAA 369
              ||||||| || ||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        374 ATAGTAGCCCAGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC-AA 432

Qy        370 AAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGT 429
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 AAGTGCCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGT 492

Qy        430 GAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCC 489
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        493 GAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCA-GCCTCTGTGGCC 551

Qy        490 ACCCCCAGGAAT-GCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTT 548
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 ACCCCCAGGAATGGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTT 611

Qy        549 CCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACT 608
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 CCTGGGCATGCCTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACT 671

Qy        609 GACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGT 668
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        672 GACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAACATTGTGT 731

Qy        669 ATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGT 728
              |||||||||||||||||||||||||||||  ||| |||| |  |  ||     |||||
Db        732 ATTGCTACTGAAAAATGGCCAGTGGAATGACTCCCCTTGCTTCCACCTCCCATCTGCCGT 791

Qy        729 CTGTGAGTTCCCTATCTGA 747
              ||||||||||||||||||
Db        792 CTGTGAGTTCCCTATCTGA 810
```

```
RESULT 15
MACMBPC
LOCUS       MACMBPC                  805 bp    mRNA    linear   PRI 26-FEB-1997
DEFINITION  Macaca mulatta mannose-binding protein C mRNA, 3' end of cds.
ACCESSION   L43911
VERSION     L43911.1  GI:1449043
KEYWORDS    mannose-binding protein.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 805)
  AUTHORS   Mogues,T., Ota,T., Tauber,A.I. and Sastry,K.N.
  TITLE     Characterization of two mannose-binding protein cDNAs from rhesus
            monkey (Macaca mulatta): structure and evolutionary implications
  JOURNAL   Glycobiology 6 (5), 543-550 (1996)
  MEDLINE   97031450
   PUBMED   8877375
FEATURES             Location/Qualifiers
     source          1. .805
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
                     /sex="male"
                     /tissue_type="liver"
     CDS             <1. .713
                     /codon_start=3
                     /product="mannose-binding protein C"
                     /protein_id="AAB48071.1"
                     /db_xref="GI:1449044"
                     /translation="CLSFICTNSTGGPQPHADGSPEVIACNSPGINGFPGKDGRDGTK
                     GEKGEPGQGLRGLQGPPGKLGPPGNPGSSGSPGPKGQKGDPGESPDCESSLAASERKA
                     LQTEMARIKKWLTFSLGRQVGNKFFLTNGEMMTFDKVKALCARFQASVATPRNAAENR
                     AIQNLIKEEAFLGITDENTEGEFVDLTGNKLTYTNWNDGEPNNAGSNEDCVLLLKNGK
                     WNDIPCSSSHLALCEFPI"
     3'UTR           714. .805
ORIGIN
```

```
  Query Match             78.7%;  Score 587.8;  DB 9;  Length 805;
  Best Local Similarity   93.6%;  Pred. No. 2.7e-162;
  Matches  613;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy         93 CTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGG 152
              | |||| |  |||||||||||||| ||| ||||||||||||||||||||||||||||||
Db         59 CAGCCCAGAGGTGATTGCCTGTAACTCCCCAGGCATCAACGGCTTCCCAGGCAAAGATGG 118

Qy        153 GCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGG 212
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        119 GCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGTTTACAGGG 178

Qy        213 CCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAA 272
              ||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        179 CCCTCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGTCTTCTGGGTCACCAGGACCAAA 238

Qy        273 GGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGA 332
              |||||||||||||||||||| | ||||| ||| |||| ||||||||||||||||||||||
Db        239 GGGCCAAAAAGGAGACCCTGGAGAGAGTCCAGATTGTGAGAGTAGCCTGGCTGCCTCAGA 298

Qy        333 AAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGG 392
              |||||||||||||||||||||||||| |||||||||| |||||||||||||||||||||||
Db        299 AAGAAAAGCTCTGCAAACAGAAATGGCTCGTATCAAAAAATGGCTGACCTTCTCTCTGGG 358

Qy        393 CAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGT 452
              || |||||||||||||||||||||||| |||||||||| |||||||||| |||||
Db        359 CAGACAAGTTGGGAACAAGTTCTTCCTGACTAATGGTGAAATGATGACCTTTGACAAAGT 418

Qy        453 GAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAA 512
              ||||||| |||||| || ||| ||||||||||||||||||||||||||||||||||||||
Db        419 GAAGGCCCTGTGTGCCAGGTTTCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAA 478

Qy        513 TGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGAC 572
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        479 CAGAGCCATCCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAACAC 538

Qy        573 AGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGG 632
              ||||||| ||||||| |||||||||||||||| ||||||||||||||||||||||||| ||
Db        539 AGAAGGGGAGTTTGTAGATCTGACAGGAAATAAACTGACCTACACAAACTGGAACGATGG 598

Qy        633 TGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTG 692
              ||| |||||||||||||||||| ||||||| |||||||| |||||||||||||| ||||
Db        599 TGAGCCCAACAATGCTGGTTCTAATGAAGACTGTGTATTGTTACTGAAAAATGGCAAGTG 658

Qy        693 GAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATCTGA 747
              ||||||| |||||||||||| | |||||||||||||| |||| ||||||||||||||||
Db        659 GAATGACATCCCCTGCTCCTCTTCCCATCTGGCCCTCTGCGAGTTCCCTATCTGA 713
```

```
Search completed: June 21, 2004, 18:23:40
Job time : 3245 secs
```

Blank

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 18, 2004, 20:52:46 ; Search time 56.5 Seconds
(without alignments)
7471.256 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 1365
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10054536/runat_18062004_173454_13562/app_query.fasta_1.903
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054536_@CGN_1_1_81_@runat_18062004_173454_13562 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1321	96.8	248	2	AAY29485	Aay29485 Human man
2	1321	96.8	248	4	AAB36578	Aab36578 Human man
3	1321	96.8	248	5	AAU75574	Aau75574 Human man
4	1321	96.8	248	7	ADE87458	Ade87458 248 amino
5	1316	96.4	248	2	AAR45005	Aar45005 Human Man
6	1309	95.9	248	3	AAY77988	Aay77988 Human MBP
7	1228	90.0	228	2	AAY03769	Aay03769 Mannan-bi
8	1137	83.3	247	1	AAP94680	Aap94680 Amino aci
9	806.5	59.1	249	3	AAB15401	Aab15401 Pig serum
10	727	53.3	409	4	AAM79745	Aam79745 Human pro
11	474	34.7	282	6	ABR42241	Abr42241 Human sur
12	454	33.3	375	3	AAY77990	Aay77990 Human SP-
13	454	33.3	375	5	AAU76473	Aau76473 Human lun
14	454	33.3	375	6	ABR44021	Abr44021 Human sur
15	442.5	32.4	351	2	AAW18780	Aaw18780 Bovine co
16	442.5	32.4	351	2	AAW13672	Aaw13672 Bovine co
17	441	32.3	374	5	AAU76471	Aau76471 Rat lung
18	440.5	32.3	371	2	AAR75642	Aar75642 Bovine co
19	408	29.9	126	5	ABG80907	Abg80907 Rat MBP p
20	407	29.8	131	5	ABG80919	Abg80919 Rat PrMBP
21	404	29.6	248	5	AAU76476	Aau76476 Antioxida
22	402	29.5	249	7	ADD48682	Add48682 Rat Prote
23	401	29.4	335	2	AAR12222	Aar12222 Conglutin
24	398.5	29.2	248	3	AAY77989	Aay77989 Human SP-
25	398.5	29.2	248	5	AAU76468	Aau76468 Human lun
26	398.5	29.2	249	7	ADD48683	Add48683 Human Pro
27	397	29.1	116	4	AAM78761	Aam78761 Human pro
28	396.5	29.0	243	7	ADD49123	Add49123 Human NOV
29	394.5	28.9	248	1	AAP60441	Aap60441 Plasmid p
30	394.5	28.9	248	1	AAP60665	Aap60665 Sequence
31	394.5	28.9	248	1	AAP80694	Aap80694 Sequence
32	394.5	28.9	248	2	AAR04215	Aar04215 Human 32K
33	394.5	28.9	248	2	AAR06331	Aar06331 Human alv
34	394.5	28.9	248	7	ADD49127	Add49127 Human NOV
35	393.5	28.8	248	2	AAR04216	Aar04216 Human 32K
36	393.5	28.8	271	2	AAR04212	Aar04212 Human 32K
37	393.5	28.8	271	2	AAR04217	Aar04217 Human 32K
38	390.5	28.6	248	1	AAP70662	Aap70662 35kd pulm
39	389.5	28.5	259	3	AAB58135	Aab58135 Lung canc
40	387.5	28.4	248	2	AAR05091	Aar05091 Vector PS
41	386.5	28.3	248	2	AAR05092	Aar05092 Product o
42	385.5	28.2	248	1	AAP60442	Aap60442 Plasmid p
43	385.5	28.2	248	1	AAP60666	Aap60666 Genomic s
44	385.5	28.2	248	1	AAP70663	Aap70663 35kd pulm
45	385.5	28.2	271	1	AAP60661	Aap60661 Genomic s

ALIGNMENTS

RESULT 1
AAY29485
ID AAY29485 standard; protein; 248 AA.
XX
AC AAY29485;
XX
DT 11-OCT-1999 (first entry)
XX
DE Human mannan-binding protein.
XX
KW Human; mannan-binding protein; hMBP; recombinant; inhibition; infection;
KW rhMBP; haemagglutination; influenza; HIV.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1. .20
FT /label= signal
FT Protein 21. .248
FT /label= mannan-binding_protein
XX
PN WO9937676-A1.
XX
PD 29-JUL-1999.
XX
PF 23-JUL-1998; 98WO-JP003311.
XX
PR 23-JAN-1998; 98JP-00011864.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
DR WPI; 1999-469114/39.

```
DR   N-PSDB; AAZ07142, AAZ07143.
XX
PT   Recombinant human mannan-binding protein expressed using pNOW1 vector.
XX
PS   Disclosure; Page 86-87; 91pp; Japanese.
XX
CC   The present invention describes recombinant human mannan-binding protein
CC   (rhMBP) having a molecular weight range of 1000-1300 or 200-400 kDa (by
CC   gel filtration with detection at 280 nm). rhMBP may be used as a
CC   component of drug compositions for the inhibition of haemagglutination
CC   and prevention of infection by viruses such as influenza and HIV. The
CC   present sequence represents hMBP (human mannan-binding protein)
XX
SQ   Sequence 248 AA;

Alignment Scores:
Pred. No.:                5.48e-115      Length:          248
Score:                    1321.00        Matches:         248
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              96.78%         Indels:          0
DB:                       2              Gaps:            0

US-10-054-536-2 (1-747) x AAY29485 (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248

RESULT 2
AAB36578
ID   AAB36578 standard; protein; 248 AA.
XX
AC   AAB36578;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human mannan binding lectin protein SEQ ID NO:1.
XX
KW   Human; mannan binding lectin; MBL; immune system; immunosuppressive;
KW   infection; cytostatic; anti-HIV; neuroprotective; anticonvulsant;
KW   antirheumatic; antiarthritic; dermatological; immunosuppressive;
KW   antiinflammatory; vasotropic; antiallergic; antiasthmatic;
KW   antiarteriosclerotic; immunostimulant; antipsoriatic; antibacterial;
KW   antidiabetic; immunomodulator; cancer; immune disorder; chemotherapy;
KW   reproductive system disorder; miscarriage; neutropaenia; HIV; epilepsy;
KW   human immunodeficiency virus; multiple sclerosis; myasthenia gravis;
KW   chronic inflammatory demyelinating polyneuropathy; rheumatoid arthritis;
KW   Eaton-Lambert's syndrome; systemic lupus erythematosus; vasculitis;
KW   Wegner's granulomatosis; Sjogren's syndrome; autoimmune neutropaenia;
KW   Crohn's disease; colitis ulcerous; asthma; septic shock syndrome;
KW   psoriasis; toxic shock syndrome; diabetes; sinusitis endocarditis;
KW   atherosclerosis; immunodeficiency; Wiskott-Aldrich syndrome; SCID;
KW   Kawasaki's disease; Guillain-Barre's syndrome.
XX
OS   Homo sapiens.
XX
PN   WO200070043-A1.
XX
PD   23-NOV-2000.
XX
PF   10-MAY-2000; 2000WO-DK000246.
XX
PR   14-MAY-1999;   99DK-00000668.
PR   20-OCT-1999;   99DK-00001508.
XX
PA   (THIE/) THIEL S.
PA   (JENS/) JENSENIUS J C.
PA   (JENS/) JENSEN T V.
XX
PI   Thiel S, Jensenius JC, Jensen TV;
XX
DR   WPI; 2001-041013/05.
XX
PT   Production of human recombinant mannan binding lectin composition for
PT   treating disorders associated with chemotherapy, HIV, by transforming
PT   host cell culture with gene expression construct and cultivating culture.
XX
PS   Disclosure; Page 60-61; 65pp; English.
XX
CC   The present invention describes a method (I) for the production of a
CC   human recombinant mannan binding lectin (rMBL) composition (C) having
CC   size distribution profile at least 50% identical to wild type MBL. The
CC   method involves preparing a gene expression construct (EC) encoding MBL
CC   or its functional equivalent, transforming a host cell culture (HC) with
CC   EC, cultivating HC, subjecting the culture medium to affinity
CC   chromatography and obtaining eluate comprising (C). (C) can have
CC   cytostatic, anti-HIV, neuroprotective, anticonvulsant, antirheumatic,
CC   antiarthritic, dermatological, immunosuppressive, antiinflammatory,
CC   vasotropic, antiallergic, antiasthmatic, antiarteriosclerotic,
CC   immunostimulant, antipsoriatic, antibacterial, antidiabetic and
CC   immunomodulator activities, and is an activator of complement cascade. A
CC   pharmaceutical composition comprising (C) can be used for the treatment
CC   of infections, MBL deficiency, cancer, immune and reproductive system
CC   disorders, disorders associated with chemotherapy, miscarriages,
CC   disorders associated with neutropaenia and human immunodeficiency virus
CC   (HIV). (C) is useful for the treatment of conditions of diseases and
CC   disorders in connection with immunosuppressive chemotherapy including, in
```

```
CC   particular infections which are seen in connection with implantation
CC   and/or transplantation of organs and diseases such as chronic
CC   inflammatory demyelinating polyneuropathy, multiple sclerosis, epilepsy,
CC   myasthenia gravis, Eaton-Lambert's syndrome, rheumatoid arthritis,
CC   systemic lupus erythematosus, vasculitis, Wegner's granulomatosis,
CC   Sjogren's syndrome, autoimmune neutropaenia, Crohn's disease, colitis
CC   ulcerous, asthma, septic shock syndrome, psoriasis, toxic shock syndrome,
CC   diabetes, sinusitis endocarditis, atherosclerosis, primary
CC   hypo/agammaglobulinaemia including common variable immunodeficiency,
CC   Wiskott-Aldrich syndrome and severe combined immunodeficiency (SCID),
CC   Kawasaki's disease, and Guillain-Barre's syndrome. The present sequence
CC   represents human MBL, which is given in the exemplification of the
CC   present invention
XX
SQ   Sequence 248 AA;

Alignment Scores:
Pred. No.:                5.48e-115      Length:          248
Score:                    1321.00        Matches:         248
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              96.78%         Indels:          0
DB:                       4              Gaps:            0

US-10-054-536-2 (1-747) x AAB36578 (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              ||||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248

RESULT 3
AAU75574
ID   AAU75574 standard; protein; 248 AA.
XX
AC   AAU75574;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human mannose-binding protein (MBP).
XX
KW   Human; mannose-binding protein; MBP; methylotrophic yeast strain;
KW   protein disulphide isomerase; PDI; heat shock protein 47; hsp47;
KW   propyl-4-hydroxylase; P4H; pathogenic organism disposal; biocide.
XX
OS   Homo sapiens.
XX
PN   US6337193-B1.
XX
PD   08-JAN-2002.
XX
PF   24-NOV-1998;   98US-00198603.
XX
PR   24-NOV-1998;   98US-00198603.
XX
PA   (APTA-) APTAGEN INC.
XX
PI   Tully RE,  Caltagirone GT,  Moyer SS,  Ronning MT;
XX
DR   WPI; 2002-163238/21.
DR   N-PSDB; ABK14771.
XX
PT   Methylotrophic yeast strain, useful for producing mannose-binding
PT   protein, comprises DNA molecules encoding the protein, protein disulfide
PT   isomerase, heat shock protein 47 and propyl-4-hydroxylase.
XX
PS   Example 1; Fig 1; 30pp; English.
XX
CC   The invention relates to a methylotrophic yeast strain (I) comprising a
CC   DNA molecule encoding mannose-binding protein (MBP), protein disulphide
CC   isomerase (PDI), heat shock protein 47 (hsp47), and propyl-4-hydroxylase
CC   (P4H), where upon culturing (I) produces the MBP, PDI, hsp47 and P4H. (I)
CC   is useful for producing MBP comprising culturing (I) under conditions
CC   suitable for the secretion of MBP by the yeast where MBP is utilised in
CC   disposal of pathogenic organisms by opsonising pathogen or activating
CC   complement cascade. The methods utilising (I) result in high yields of
CC   MBP without the use of foetal calf serum, in a cost-effective manner. The
CC   present sequence relates to the amino acid sequence of human mannose-
CC   binding protein
XX
SQ   Sequence 248 AA;

Alignment Scores:
Pred. No.:                5.48e-115      Length:          248
Score:                    1321.00        Matches:         248
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              96.78%         Indels:          0
DB:                       5              Gaps:            0

US-10-054-536-2 (1-747) x AAU75574 (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40
```

```
Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              ||||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248
```

RESULT 4
ADE87458
ID ADE87458 standard; protein; 248 AA.
XX
AC ADE87458;
XX
DT 29-JAN-2004 (first entry)
XX
DE 248 amino acid human mannan binding lectin protein.
XX
KW mannan binding lectin; MBL; MBL-associated serine protease; MASP; MASP-1;
KW MASP-2; MASP-3; antibacterial; immunostimulant; immunocompromised.
XX
OS Homo sapiens.
XX
PN EP1344533-A1.
XX
PD 17-SEP-2003.
XX
PF 14-MAR-2003; 2003EP-00388016.
XX
PR 15-MAR-2002; 2002DK-00000414.
XX
PA (NATL-) NATLMMUNE AS.
XX
PI Larsen JL, Kongerslev L;
XX
DR WPI; 2003-758222/72.
XX
PT A pharmaceutical composition containing mannan binding lectin or its
PT variant is useful to treat clinical conditions, particularly infection,
PT and immunosuppressive conditions.
XX
PS Claim 18; SEQ ID NO 1; 44pp; English.
XX
CC The invention relates to a novel pharmaceutical composition comprising
CC additives and at least 200 micrograms/ml protein containing material
CC where mannan binding lectin (MBL) and/or MBL variant constitute at least
CC 35% (w/w) of total protein, or comprising at least 400 micrograms/ml MBL
CC and/or MBL variant. The protein preferably further comprises one or more
CC MBL-associated serine protease(s) (MASP) or its fragments, particularly
CC MASP-1, MASP-2 or MASP-3 having respectively the 699, 686 or 728 amino
CC acid sequence fully defined in the specification. The MBL preferably has
CC the 248 sequence fully defined in the specification or it's fragment and
CC may be purified naturally occurring human serum MBL or recombinant MBL.
CC The novel pharmaceutical compositions have antibacterial and
CC immunostimulant activities. The composition can be used to treat a
CC clinical condition, particularly an infection, especially where the
CC individual has a subnormal MBL level. The composition may be used to
CC treat immunocompromised conditions. This sequence represents the 248
CC amino acid mannan binding lectin protein of the invention.
XX
SQ Sequence 248 AA;

Alignment Scores:

Pred. No.:	5.48e-115	Length:	248
Score:	1321.00	Matches:	248
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.78%	Indels:	0
DB:	7	Gaps:	0

US-10-054-536-2 (1-747) x ADE87458 (1-248)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248

RESULT 5
AAR45005
ID   AAR45005 standard; protein; 248 AA.
XX
AC   AAR45005;
XX
DT   25-MAR-2003  (revised)
DT   13-JUN-1994  (first entry)
XX
DE   Human Mannose-Binding Protein.
XX
KW   human MBP; N-acetylglucosamine; fucose; mannose; lectin;
KW   carbohydrate binding domain; complement fixing domain; anti-viral agent.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .62
FT                   /note= "Cysteine-rich region, consistent with formation
FT                   of multimers in the native mol."
FT   Domain          63. .101
FT                   /label= complement-fixing_domain
FT                   /note= "collagen-like segment with Gly-X-Y repeats"
FT   Misc-difference 75
FT                   /note= "corresponds to AAC codon in AAQ53529"
FT   Misc-difference 77
FT                   /note= "corresponds to GCC codon in AAQ53529"
FT   Domain          102. .124
FT                   /label= Cell_attachment_domain
FT                   /note= "putative; facilitates attachment and ingestion by
FT                   phagocyte cells"
FT   Domain          125. .248
FT                   /label= Carbohydrate-binding_domain
FT                   /note= "mannose-binding domain is within this region"
FT   Misc-difference 197
FT                   /note= "corresponds to GAT codon in AAQ53529"
XX
PN   US5270199-A.
XX
PD   14-DEC-1993.
XX
PF   06-FEB-1992;   92US-00831619.
XX
PR   20-AUG-1987;   87US-00087628.
PR   05-OCT-1989;   89US-00417822.
XX
PA   (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI   Ezekowitz RAB;
XX
DR   WPI; 1993-404934/50.
DR   N-PSDB; AAQ53529, AAQ53530.
XX
PT   Recombinant human mannose-binding protein polypeptide(s) - used in
PT   diagnosis or therapy of infections caused by bacteria, fungi or viruses,
PT   partic. HIV.
XX
PS   Claim 3; Fig 2; 14pp; English.
XX
CC   MBP is a soluble lectin-like molecule which is synthesised in hepatocytes
CC   and released into the bloodstream. Human MBP binds carbohydrate and can
CC   selectively recognise configurations of high mannose, fucose or N-
CC   acetylglucosamine which are present on pathogens. Interaction of MBP with
CC   mannose-rich pathogens enhances clearance of the pathogens by
CC   phagocytosis or results in activation of complement. Sequences (cDNA and
CC   genomic DNA) coding for human MBP have been isolated. (Updated on 25-MAR-
CC   2003 to correct PF field.)
XX
SQ   Sequence 248 AA;

Alignment Scores:
Pred. No.:                1.61e-114      Length:          248
Score:                    1316.00        Matches:         247
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    99.60%         Mismatches:      0
Query Match:              96.41%         Indels:          0
DB:                       2              Gaps:            0

US-10-054-536-2 (1-747) x AAR45005 (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAsnLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||||||||||||||||||||||
Db        241 LeuAlaValCysGluPheProIle 248
```

```
RESULT 6
AAY77988
ID   AAY77988 standard; protein; 248 AA.
XX
AC   AAY77988;
XX
DT   20-JUN-2000  (first entry)
XX
DE   Human MBP amino acid sequence.
XX
KW   Collectin; human; antibacterial; antiviral; MBP.
XX
OS   Homo sapiens.
XX
PN   WO200011161-A1.
XX
PD   02-MAR-2000.
XX
PF   24-AUG-1999;   99WO-JP004552.
XX
PR   24-AUG-1998;   98JP-00237611.
XX
PA   (FUSO ) FUSO PHARM IND LTD.
XX
PI   Wakamiya N;
XX
DR   WPI; 2000-224696/19.
XX
PT   New collectin of human origin having antibacterial and antiviral
PT   activity, and gene encoding it useful for production of transgenic
PT   animals and of antibodies for screening potential drug molecules.
XX
PS   Disclosure; Fig 2-3; 106pp; Japanese.
XX
CC   The invention relates to polynucleotides encoding a new collectin of
CC   human origin. The collectin can be used as an antibacterial and antiviral
CC   agent and for screening potential drug molecules. The new collectin can
CC   be produced by standard recombinant methodology. The present sequence
CC   represents a human MBP, used in homology studies with collectin
XX
SQ   Sequence 248 AA;

Alignment Scores:
Pred. No.:                7.3e-114        Length:          248
Score:                    1309.00         Matches:         246
Percent Similarity:       99.19%          Conservative:    0
Best Local Similarity:    99.19%          Mismatches:      2
Query Match:              95.90%          Indels:          0
DB:                       3               Gaps:            0

US-10-054-536-2 (1-747) x AAY77988 (1-248)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 40

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 60

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 80

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerIleSerHis 240

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              |||   ||||||||||||||||||
Db        241 LeuAsnValCysGluPheProIle 248

RESULT 7
AAY03769
ID   AAY03769 standard; protein; 228 AA.
XX
AC   AAY03769;
XX
DT   11-JUN-1999  (first entry)
XX
DE   Mannan-binding protein (MBP).
XX
KW   Mannan-binding protein; MBP; carcinostatic; anticancer agent; tumour;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   WO9910001-A1.
XX
PD   04-MAR-1999.
XX
PF   19-AUG-1998;   98WO-JP003697.
XX
PR   21-AUG-1997;   97JP-00239113.
XX
PA   (TAKI ) TAKARA SHUZO CO LTD.
XX
PI   Kawasaki T;
XX
DR   WPI; 1999-190466/16.
DR   N-PSDB; AAX29295.
XX
PT   Carcinostatic agent containing mannan-binding protein - inhibits cancer
PT   growth and proliferation into surrounding tissues.
XX
PS   Claim 2; Page 12-13; 23pp; Japanese.
XX
CC   The invention relates to cancer inhibiting agents that contain an active
CC   component, mannan-binding protein (or the gene encoding it). Mannan-
CC   binding protein (MBP) has carcinostatic activity, and can be used as the
CC   active component in anticancer agents for inhibiting tumour growth and
CC   suppressing the proliferation of a cancer into surrounding tissue.
CC   Alternatively, the gene encoding MBP may be delivered to the tumour in a
```

```
CC   plasmid or viral vector (especially in a vaccinia or adenovirus vector).
CC   The MBP may be human MBP, or derived from it by addition, deletion and/or
CC   substitution of one or more amino acid residues. The agents containing
CC   the MBP can be used for effective treatment of malignant tumours
CC   (including solid tumours) by direct application to the tumour. The
CC   present sequence represents a MBP sequence
XX
SQ   Sequence 228 AA;

Alignment Scores:
Pred. No.:                2.79e-106      Length:          228
Score:                    1228.00        Matches:         228
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              89.96%         Indels:          0
DB:                       2              Gaps:            0

US-10-054-536-2 (1-747) x AAY03769 (1-228)

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer 20

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly 40

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly 60

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 80

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 100

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 120

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 140

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 160

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 180

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 200

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspCysValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 220

Qy        721 CTGGCCGTCTGTGAGTTCCCTATC 744
              ||||||||||||||||||||||||
Db        221 LeuAlaValCysGluPheProIle 228

RESULT 8
AAP94680
ID   AAP94680 standard; protein; 247 AA.
XX
AC   AAP94680;
XX
DT   25-MAR-2003  (revised)
DT   29-JUN-1990  (first entry)
XX
DE   Amino acid sequence of human mannose binding protein (hMBP) as encoded by
DE   liver cDNA.
XX
KW   Human mannose binding protein; liver cDNA.
XX
OS   Homo sapiens.
XX
PN   WO8901519-A.
XX
PD   23-FEB-1989.
XX
PF   05-AUG-1988;   88WO-US002591.
XX
PR   20-AUG-1987;   87US-00087628.
XX
PA   (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.
XX
PI   Ezekowitz RAB;
XX
DR   WPI; 1989-068874/09.
DR   N-PSDB; AAN91079.
XX
PT   New DNA encoding human mannose binding protein or fragment - for treating
PT   infections, used for deriving diagnostic antibodies.
XX
PS   Disclosure; Fig 2; 34pp; English.
XX
CC   The patent claims cDNA encoding 20 or more contiguous amino acids of
CC   hMBP, (expression) vectors and cells contg. it, peptides encoded by it,
CC   and antibodies (Ab) against these peptides. Peptides encoded by it, opt.
CC   fused to a toxin, are used to treat animals infected with bacterial,
CC   fungal or viral (including HIV) pathogens. They bind to mannose units on
CC   such organisms and cause host defence cells to be attracted to them. The
CC   peptides inhibit growth or infection (esp. in HIV, where concn. is 1-500
CC   microgram/ml. final serum concn. in human) of the organism. The
CC   antibodies are useful in diagnosing susceptibility to infection (by
CC   ELISA). (Updated on 25-MAR-2003 to correct PD field.) (Updated on 25-MAR-
CC   2003 to correct PA field.)
XX
SQ   Sequence 247 AA;

Alignment Scores:
Pred. No.:                9.65e-98       Length:          247
Score:                    1137.00        Matches:         229
Percent Similarity:       92.37%         Conservative:    1
Best Local Similarity:    91.97%         Mismatches:      15
Query Match:              83.30%         Indels:          6
DB:                       1              Gaps:            0

US-10-054-536-2 (1-747) x AAP94680 (1-247)

Qy          8 TGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTG 67
              |||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db          3 CysPhe-IleThrProSer-LeuLeuLeuSerMetValAlaAlaSerTyrSerGluThrV 22

Qy         68 TGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCA 127
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db         22 alThrCysGluGlyAlaGlnLysThrCysProAlaValIleAlaCysSerSerProGlyI 42

Qy        128 TCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAG 187
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||      |
Db         42 leAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyArgLysGlyGlyThrG 62

Qy        188 GCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAG 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 lyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGlyAsnProG 82

Qy        248 GGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATG 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 lyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSerProAspG 102
```

```
Qy        308 GTGATAGTAG-CCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
              ||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 lyAspSerSerProGlyCysLeuArgLysLysSerSerAlaAsnArgAsnGlyThrTyrG 122

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              |||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 lnLysCys-LeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeuThrAsn 141

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        142 GlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGln-ProLeuTr 161

Qy        487 GCCACCCCCAGGAATGCT-GCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGC 545
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        161 pProProProGlyMetAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGluGluAl 181

Qy        546 CTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAG 605
              ||||||||||:::   ||||||||||||||||||||||||||||||||||||||||||||
Db        181 aPheLeuGlyMetProAspGluLysThrGluGlyGlnPheValAspLeuThrGlyAsnAr 201

Qy        606 ACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTG 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||
Db        201 gLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGluHisCy 221

Qy        666 TGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGC 725
              |||||||||||||||||||||||||||||||   ||||||                ||
Db        221 sValLeuLeuLeuLysAsnGlyGlnTrpAsnAspSerProCysPheHisLeuProSerAl 241

Qy        726 CGTCTGTGAGTTCCCTATC 744
              ||||||||||||||||||
Db        241 aValCysGluPheProIle 247

RESULT 9
AAB15401
ID   AAB15401 standard; protein; 249 AA.
XX
AC   AAB15401;
XX
DT   17-JAN-2001  (first entry)
XX
DE   Pig serum lectin.
XX
KW   Pig; serum; lectin; phosphomannan affinity; Hansunela capsulata;
KW   physiologically active substance.
XX
OS   Sus scrofa.
XX
PN   JP2000166569-A.
XX
PD   20-JUN-2000.
XX
PF   09-DEC-1998;   98JP-00350283.
XX
PR   09-DEC-1998;   98JP-00350283.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2000-485555/43.
DR   N-PSDB; AAA70738.
XX
PT   New pig serum lectin gene encoding phosphomanna.
XX
PS   Claim 3; Fig 1; 12pp; Japanese.
XX
CC   The invention relates to the isolation of a novel pig serum lectin gene
CC   which encodes a protein with affinity for a phosphomannan. The protein
CC   was isolated on a phosphomannan affinity column generated by coupling
CC   phosphomannan from cultured Hansunela capsulata to an epoxy column. The
CC   pig lectin can be used as a new physiologically active substance
XX
SQ   Sequence 249 AA;
```

```
Alignment Scores:
Pred. No.:              9.19e-67        Length:         249
Score:                  806.50          Matches:        154
Percent Similarity:     75.50%          Conservative:   34
Best Local Similarity:  61.85%          Mismatches:     58
Query Match:            59.08%          Indels:         3
DB:                     3               Gaps:           2

US-10-054-536-2 (1-747) x AAB15401 (1-249)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||   ||||||   |||||||||:::|||||||||   :::|||   ||||||||||||
Db          1 MetLeuLeuPheSerSerLeuProValLeuLeuLeuCysValValThrAlaSerTyrSer 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              |||      ||||||||||||||||||||||||   :::||||||   |||
Db         21 GluIleLysThrCysGluAspAlaGlnLysThrCys---SerValIleThrCysGlyIle 39

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||      ||||||   ||||||:::|||||||||||||||   ||||||||||||||
Db         40 ProValThrAsnGlyThrProGlyArgAspGlyArgAspGlyArgLysGlyGluLysGly 59

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              |||||||||   ||||||||||||   |||||||||||||||||:::|||||||||||||
Db         60 GluProGlyProGlyLeuArgGlySerGlnGlyProProGlyLysMetGlyProProGly 79

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              |||   |||   :::|||       |||||||||||||||||||||||||||   :::
Db         80 AsnIleGlyGluThrGlyProLeuGlyProLysGlyGlnLysGlyAspProGlyAspThr 99

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGA------AAAGCTCTGCAAACAGAA 354
                  ::::::   ||||||      |||          :::   |||::::::|||
Db        100 SerGlyValGluAlaLysLeuAlaAsnLeuGluGlyGlnIleArgIleLeuLysSerGlu 119

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
                  :::      :::||||||      |||||||||||||||||:::         |||
Db        120 LeuAspHisValLysLysLeuGlnThrPheSerLeuGlyLysLysSerArgLysLysLeu 139

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
                  ::::::||||||||||||:::|||   |||   ||||||||   ||||||   :::
Db        140 TyrValThrAsnGlyGluMetMetProPheSerLysValLysThrLeuCysAlaGluLeu 159

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
              ||||||:::||||||||||||:::||||||   ||||||   ||||||||::::::
Db        160 GlnAlaThrValAlaThrProLysAsnAlaGluGluAsnLysAlaIleGlnAspMetAla 179

Qy        535 AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTG 594
                 :::   ||||||||||||||||||||||   |||||||||||||:::   :::
Db        180 ProAspValAlaPheLeuGlyIleThrAspGluValThrGluGlyGlnPheMetTyrVal 199

Qy        595 ACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCT 654
              ||||||   |||:::||||||:::||||||         |||||||||:::   ||||||
Db        200 ThrGlyGlyArgMetThrTyrSerAsnTrpLysSerAsnGluProAsnAspHisGlySer 219

Qy        655 GATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACC 714
                 ||||||||||||:::|||   ::::::|||   |||||||||:::   ||||||:::
Db        220 GlyGluAspCysValIleLeuGlnArgAspGlyLeuTrpAsnAspIleSerCysSerSer 239

Qy        715 TCCCATCTGGCCGTCTGTGAGTTCCCT 741
              |||   ||||||||||||||||||||
Db        240 SerPheLeuAlaValCysGluPhePro 248

RESULT 10
AAM79745
ID   AAM79745 standard; protein; 409 AA.
XX
AC   AAM79745;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3391.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
```

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52878.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 20; Page 321; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 409 AA;

Alignment Scores:

Pred. No.:	2.96e-59	Length:	409
Score:	727.00	Matches:	155
Percent Similarity:	46.97%	Conservative:	31
Best Local Similarity:	39.14%	Mismatches:	60
Query Match:	53.26%	Indels:	150
DB:	4	Gaps:	3

US-10-054-536-2 (1-747) x AAM79745 (1-409)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||   |||||||||||||   |||:::|||||||||||||:::|||   ||||||   |||
Db         14 MetPheLeuPheProSerPheProValLeuLeuLeuSerValValThrAlaSerCysSer 33

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              :::|||      |||   |||   ||||||||||||   ::::::|||   |||
Db         34 LysThrLysAlaCysAlaAspThrGlnLysThrCys---SerMetIleThrCysGlyIle 52

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              |||      ||||||   ||||||:::||||||||||||      ||||||||||||||
Db         53 ProValThrAsnGlyThrProGlyArgAspGlyArgAspArgProLysGlyGluLysGly 72

Qy        181 GAACCA---------------------------------------------------- 186
              ||||||
Db         73 GluProGlyLeuGlyGlnValSerValAlaSer***IleSerThrSerGlyArgCysSer 92

Qy        186 ---------------------------------------------------------- 186

Db         93 SerLysSerValLeuGluProAlaThrArgGlyLeuLysHisArgLeuGlyGluAlaPro 112

Qy        186 ---------------------------------------------------------- 186

Db        113 LeuSerSerGlyProMetLeuHisSerGluGlnProLeu***AsnAlaIleAlaSerLys 132

Qy        186 ---------------------------------------------------------- 186

Db        133 ThrLysLeuPheValAspSerLeuGlySerHisIleSerThrGlnGluLeuGlyValCys 152

Qy        186 ---------------------------------------------------------- 186

Db        153 GlyCysProPheArgGlyValSerCysLeuValGlyGluLeuAlaLeuValGlnAlaLeu 172

Qy        186 ---------------------------------------------------------- 186

Db        173 His***ValAlaGlyGluSerPhePhePheGlySerAspHisTrpLeuIleGlyCysAla 192

Qy        186 ---------------------------------------------------------- 186

Db        193 GlyGlyGluGlnGluTrpSerIleGluLeuLeuGlyLysLysLysArgValThrAlaThr 212

Qy        187 ----------------------------GGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCT 219
                                          ||||||||||||||||||||||||||||||
Db        213 GlySerSerSerLeuCysLeuAlaThrGlyGlnGlyLeuArgGlyLeuGlnGlyProPro 232

Qy        220 GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAA 279
              ||||||:::|||||||||||||||   |||   ||||||   |||||||||:::||||||
Db        233 GlyLysMetGlyProProGlyAsnThrGlyThrSerGlyIleProGlyProArgGlyGln 252

Qy        280 AAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAA 339
              ||||||||   |||   :::         ::::::   ||||||      |||||||||
Db        253 LysGlyAspArgGlyAspAsnSerValAlaGluAlaLysLeuAlaAsnLeuGluArgLys 272

Qy        340 ------GCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGC 393
                  :::|||::::::|||:::         ||||||         ||||||||||||
Db        273 Leu***SerLeuArgSerGluLeuAspHisThrLysLysLeu***ProPheSerLeuGly 292

Qy        394 AAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTG 453
              |||      |||   |||   |||:::||||||||||||   |||   |||   ||||||
Db        293 LysMetSerGlyLysLysLeuPheValThrAsnGlyGluArgMetProPheSerLysVal 312

Qy        454 AAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAAT 513
              ||||||||||||         ||||||:::||||||   |||:::||||||   ||||||
Db        313 LysAlaLeuCysAlaGlyLeuGlnAlaThrValAlaAlaProLysAsnAlaGluGluAsn 332

Qy        514 GGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACA 573
                 |||||||||::::::   |||:::   ||||||||||||||||||||||||   |||
Db        333 LysAlaIleGlnAspValAlaLysAspThrAlaPheLeuGlyIleThrAspGluAlaThr 352

Qy        574 GAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGT 633
              ||||||||||||:::   |||||||||   ||||||||||||:::||||||   :::
Db        353 GluGlyGlnPheMetTyrLeuThrGlyGlyArgLeuThrTyrSerAsnTrpLysLysAsp 372

Qy        634 GAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGG 693
              |||||||||:::   ||||||   ||||||||||||:::||||||   ||||||   |||
Db        373 GluProAsnAspHisGlySerGlyGluAspCysValIleLeuLeuAsnAsnGlyLeuTrp 392

Qy        694 AATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCT 741
              |||   :::   |||::::::|||   :::|||:::||||||||||||
Db        393 AsnGlyIleSerCysThrSerSerPheIleAlaIleCysGluPhePro 408
```

RESULT 11
ABR42241
ID ABR42241 standard; protein; 282 AA.
XX
AC ABR42241;

```
XX
DT   28-JUL-2003  (first entry)
XX
DE   Human surfactant protein-D (endothelial form).
XX
KW   Human; surfactant protein-D; SP-D; atherosclerosis; antiarteriosclerotic;
KW   antidiabetic; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /label= Leader
FT   Region          22. .45
FT                   /label= N-terminal
FT   Region          46. .129
FT                   /label= Collagen
FT   Modified-site   90
FT                   /note= "N-glycosylated"
FT   Region          130. .157
FT                   /label= Neck
FT   Domain          158. .282
FT                   /label= Carbohydrate_recognition
XX
PN   WO2003035683-A2.
XX
PD   01-MAY-2003.
XX
PF   25-OCT-2002; 2002WO-DK000711.
XX
PR   26-OCT-2001; 2001US-0330664P.
PR   27-DEC-2001; 2001US-0342372P.
XX
PA   (HOLM/) HOLMSKOV U.
PA   (LYST/) LYSTER G L.
PA   (MADS/) MADSEN J.
PA   (TORN/) TORNOE I.
XX
PI   Holmskov U,  Lyster GL,  Madsen J,  Tornoe I;
XX
DR   WPI; 2003-421403/39.
DR   N-PSDB; ACC57755.
XX
PT   New nucleotide sequence encoding surfactant protein-D, useful for
PT   preparing a composition for preventing or treating diabetes or
PT   atherosclerosis.
XX
PS   Claim 3; Fig 1; 78pp; English.
XX
CC   The present sequence is the protein sequence of a novel form of human
CC   surfactant protein D (SP-D) that is expressed in endothelial cells,
CC   inhibits oxidation of low density lipoprotein and protects endothelial
CC   cells from oxidative damage. The endothelial form polypeptide lacks 93
CC   amino acids compared to the lung form of SP-D and has only 28 Gly-Xaa-Xaa
CC   repeats as compared to 57 in the lung form of SP-D. The invention
CC   provides a novel method of diagnosing a person's susceptibility for
CC   having an increased risk for the development of atherosclerosis by
CC   measuring the amount of SP-D in plasma and serum samples. An increased
CC   risk for the development of atherosclerosis in a human is considered at a
CC   detected serum or plasma SP-D concentration below 1104.6 ng/ml, in
CC   particular below 743 ng/ml, and especially below 370 ng/ml or 152.7
CC   ng/ml. SP-D, or a nucleic acid encoding it, can be used to prevent or
CC   treat an atherosclerosis-related disease such as stroke, kidney failure,
CC   blindness, leg amputation and myocardial infarction, and to treat obesity
CC   or diabetes
XX
SQ   Sequence 282 AA;

Alignment Scores:
Pred. No.:                1.39e-35        Length:          282
Score:                    474.00          Matches:         114
Percent Similarity:       52.65%          Conservative:    35
Best Local Similarity:    40.28%          Mismatches:      96

Query Match:              34.73%          Indels:          38
DB:                       6               Gaps:            9

US-10-054-536-2 (1-747) x ABR42241 (1-282)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||   ||||||          |||:::||||||:::             |||
Db          1 MetLeuLeuPheLeuLeuSerAlaLeuValLeuLeuThrGln---ProLeuGlyTyrLeu 19

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCA---------GTGATTGCC 111
              |||                 :::   :::|||   |||:::         ::::::
Db         20 GluAlaGluMetLysThrTyrSerHisArgThrThrProSerAlaCysThrLeuValMet 39

Qy        112 TGTAGCTCT------CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACC 165
              |||||||||      |||:::  |||      |||:::|||||||||:::|||
Db         40 CysSerSerValGluSerGlyLeuProGlyArgAspGlyArgAspGlyArgGluGlyPro 59

Qy        166 AAGGGAGAAAAGGGGGAACCAGGC--------------------CAAGGGCTCAGAGGC 204
              :::|||||||||||:::||||||                    |||:::   |||
Db         60 ArgGlyGluLysGlyAspProGlyLeuProGlyAlaAlaGlyGlnAlaGlyMetProGly 79

Qy        205 TTACAGGGCCCC-------------------------CCTGGAAAGTTGGGGCCTCCA 237
              ||||||                             |||:::  ||||||
Db         80 GlnAlaGlyProValGlyProLysGlyAspAsnGlySerValGlyGluProGlyProLys 99

Qy        238 GGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA--- 294
              |||:::  ||||||  |||  ||||||  ||||||  ||||||  ||||||
Db        100 GlyAspThrGlyProArgGlyProProGlyLeuLysGlyAspLysGlyIleProGlyAsp 119

Qy        295 AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGAAAAGCT------ 342
              |||         |||:::|||   |||        ||||||  |||:::
Db        120 LysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArgGlnGlnValGlu 139

Qy        343 ----------------------CTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTG 378
                                   ||||||        ::::::  ||||||
Db        140 AlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyrLysLysValGlu 159

Qy        379 ACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATG 438
              |||      |||:::  ||||||  |||  |||  |||  |||  :::
Db        160 LeuPheProAsnGlyGlnGlyValGlyGluLysIlePheLysThrAlaGlyPheValLys 179

Qy        439 ACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGG 498
              |||  :::  :::  ||||||  :::          :::|||:::||||||
Db        180 ProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeuAlaSerProArg 199

Qy        499 AATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAGGAGGAAGCCTTC 549
              :::||||||||||||  |||:::|||  |||:::          |||  ||||||
Db        200 SerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsnGluAlaAlaPhe 219

Qy        550 CTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTG 609
              |||  :::||||||  ||||||||||||:::|||         ||||||     |||
Db        220 LeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThrGlyGluSerLeu 239

Qy        610 ACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTA 669
              |||:::||||||      ||||||||||||:::  |||      ||||||||||||
Db        240 ValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySerGluAspCysVal 259

Qy        670 TTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTC 729
              :::      ||||||:::||||||||      |||              |||  |||
Db        260 GluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLysArgLeuValVal 279

Qy        730 TGTGAGTTC 738
              |||||||||
Db        280 CysGluPhe 282

RESULT 12
AAY77990
ID   AAY77990 standard; protein; 375 AA.
XX
AC   AAY77990;
XX
DT   20-JUN-2000  (first entry)
XX
```

DE Human SP-D amino acid sequence.
XX
KW Collectin; human; antibacterial; antiviral; SP-D.
XX
OS Homo sapiens.
XX
PN WO200011161-A1.
XX
PD 02-MAR-2000.
XX
PF 24-AUG-1999; 99WO-JP004552.
XX
PR 24-AUG-1998; 98JP-00237611.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
DR WPI; 2000-224696/19.
XX
PT New collectin of human origin having antibacterial and antiviral
PT activity, and gene encoding it useful for production of transgenic
PT animals and of antibodies for screening potential drug molecules.
XX
PS Disclosure; Fig 2-3; 106pp; Japanese.
XX
CC The invention relates to polynucleotides encoding a new collectin of
CC human origin. The collectin can be used as an antibacterial and antiviral
CC agent and for screening potential drug molecules. The new collectin can
CC be produced by standard recombinant methodology. The present sequence
CC represents a human SP-D, used in homology studies with collectin
XX
SQ Sequence 375 AA;

Alignment Scores:

Pred. No.:	1.12e-33	Length:	375
Score:	454.00	Matches:	99
Percent Similarity:	55.07%	Conservative:	26
Best Local Similarity:	43.61%	Mismatches:	82
Query Match:	33.26%	Indels:	20
DB:	3	Gaps:	5

US-10-054-536-2 (1-747) x AAY77990 (1-375)

```
Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||:::   |||      |||      |||      |||   ||||||||:::
Db        149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||      |||      |||      |||
Db        169 GlyValProGlyGluArgGlyValProGlyAsnAlaGlyAlaAlaGlySerAlaGlyAla 188

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||:::||||||      |||   ||||||   ||||||   ||||||
Db        189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy        286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
              ||||||   |||         |||:::|||   |||         ||||||   |||
Db        209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy        337 AAAGCT----------------------------CTGCAAACAGAAATGGCACGTATC 366
              :::                              ||||||        ::::::
Db        229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||      |||      |||:::   ||||||   |||   |||   |||
Db        249 LysLysValGluLeuProProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::      |||   :::   :::   ||||||   :::            :::
Db        269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
              |||:::||||||:::||||||||||||   |||:::|||   |||:::
Db        289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              |||   ||||||||||   :::||||||   ||||||||||||:::|||          |||
Db        309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
              |||      |||   |||:::||||||      ||||||||||||:::   |||
Db        329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySer 348

Qy        658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              ||||||||||||   :::      ||||||:::||||||||      |||
Db        349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy        718 CATCTGGCCGTCTGTGAGTTC 738
                |||   ||||||||||||
Db        369 ArgLeuValValCysGluPhe 375
```

RESULT 13
AAU76473
ID AAU76473 standard; protein; 375 AA.
XX
AC AAU76473;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human lung surfactant protein D.
XX
KW Antioxidant; lung surfactant protein D; antiarteriosclerotic;
KW antimicrobial; lipid oxidation; acute lung injury; atherosclerosis;
KW food spoilage; cosmetic; adult respiratory distress syndrome; human;
KW hyperoxic lung injury; butylated hydroxy toluene; BHT; thimerosal.
XX
OS Homo sapiens.
XX
PN WO200206301-A2.
XX
PD 24-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-US021226.
XX
PR 30-JUN-2000; 2000US-0215313P.
XX
PA (UYCI-) UNIV CINCINNATI.
XX
PI Mccormack FX;
XX
DR WPI; 2002-216994/27.
XX
PT Method useful for treating conditions associated with lipid oxidation or
PT acute lung injury, preventing microbial contamination and spoilage of
PT foods, cosmetics and medications, comprises antioxidant/antimicrobial
PT lung surfactant protein.
XX
PS Claim 3; Page 10; 88pp; English.
XX
CC The invention relates to a method of treating conditions associated with
CC lipid oxidation, acute lung injury, atherosclerosis, inhibiting or
CC reducing microbial contamination, preventing microbial proliferation,
CC contamination or colonisation, and preventing spoilage of foods,
CC cosmetics and medications. The method involves using an antioxidant or
CC antimicrobial lung surfactant protein compound. The method is also useful
CC for treating adult respiratory distress syndrome and hyperoxic lung
CC injury. The surfactant proteins are advantageous over antioxidant
CC ingredients such as butylated hydroxy toluene (BHT) or antimicrobial
CC ingredients such as thimerosal for the preparation of all natural
CC products, since they are derived from animal sources, are edible and are
CC non-toxic. The present sequence represents human lung surfactant protein
CC D used in the method of the invention
XX
SQ Sequence 375 AA;

```
Alignment Scores:
Pred. No.:                1.12e-33        Length:          375
Score:                    454.00          Matches:         99
Percent Similarity:       55.07%          Conservative:    26
Best Local Similarity:    43.61%          Mismatches:      82
Query Match:              33.26%          Indels:          20
DB:                       5               Gaps:            5

US-10-054-536-2 (1-747) x AAU76473 (1-375)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||:::   |||      |||      |||      |||   |||||||||:::
Db        149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||      |||      |||      |||
Db        169 GlyValProGlyGluArgGlyValProGlyAsnThrGlyAlaAlaGlySerAlaGlyAla 188

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||:::||||||      |||   ||||||  ||||||  ||||||
Db        189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy        286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
                ||||||   |||        |||:::|||   |||        ||||||   |||
Db        209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy        337 AAAGCT----------------------------CTGCAAACAGAAATGGCACGTATC 366
              :::                              ||||||        ::::::
Db        229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||       |||      |||:::   ||||||  |||   |||   |||
Db        249 LysLysValGluLeuProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::     |||   :::   :::   ||||||  :::           :::
Db        269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
              |||:::||||||:::|||||||||||||   |||:::|||   |||:::
Db        289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              |||   ||||||||   :::||||||   ||||||||||||:::|||         |||
Db        309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
              |||      |||   |||:::||||||      ||||||||||:::   |||
Db        329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySer 348

Qy        658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              ||||||||||||   :::      ||||||:::|||||||||      |||
Db        349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy        718 CATCTGGCCGTCTGTGAGTTC 738
                |||   ||||||||||||
Db        369 ArgLeuValValCysGluPhe 375

RESULT 14
ABR44021
ID   ABR44021 standard; protein; 375 AA.
XX
AC   ABR44021;
XX
DT   04-AUG-2003  (first entry)
XX
DE   Human surfactant protein D (SP-D).
XX
KW   Surfactant protein D; SP-D; antiinflammatory; antiasthmatic; rSPD(n/CRD);
KW   dermatological; antiallergic; antibacterial; virucide; fungicide; human;
KW   antiatherosclerotic; immunosuppressive.
XX
OS   Homo sapiens.
XX
PN   WO2003035679-A2.
XX
PD   01-MAY-2003.
XX
PF   25-OCT-2002; 2002WO-GB004824.
XX
PR   25-OCT-2001; 2001GB-00025638.
PR   26-APR-2002; 2002GB-00009619.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Clark H,  Nadesalingam P,  Reid KBM,  Strong P;
XX
DR   WPI; 2003-430400/40.
DR   N-PSDB; ACC48169.
XX
PT   Novel recombinant surfactant protein D, termed rSPD n/CRD polypeptide, or
PT   its fragment, homolog, variant or derivative, useful for treating
PT   inflammatory diseases, asthma, allergy, or atherosclerosis.
XX
PS   Disclosure; Page 126-127; 167pp; English.
XX
CC   The invention relates to a recombinant surfactant protein D (rSPD)(n/CRD)
CC   polypeptide (I), or its fragment, homologue, variant or derivative, for
CC   use in treatment or prophylaxis of a disease. (I) is useful for treating
CC   an individual suffering from a disease or preventing the occurrence of a
CC   disease in an individual, where the disease comprises an inflammatory
CC   disease, preferably eczema or an inflammatory lung disease (selected from
CC   neonatal chronic lung disease, neonatal respiratory distress syndrome
CC   (RDS), adult respiratory distress syndrome, chronic obstructive airways
CC   disease (COPD), asthma, cystic fibrosis, pulmonary fibrosis, emphysema,
CC   interstitial inflammatory lung disease, sarcoidosis, pneumonia, chronic
CC   inflammatory lung disease, neonatal chronic inflammatory lung disease)
CC   and allergy (see ACC48168 for a detailed description of all the diseases
CC   that can be treated using (I)). The present sequence represents a human
CC   surfactant protein (SP-D)
XX
SQ   Sequence 375 AA;

Alignment Scores:
Pred. No.:                1.12e-33        Length:          375
Score:                    454.00          Matches:         99
Percent Similarity:       55.07%          Conservative:    26
Best Local Similarity:    43.61%          Mismatches:      82
Query Match:              33.26%          Indels:          20
DB:                       6               Gaps:            5

US-10-054-536-2 (1-747) x ABR44021 (1-375)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||:::   |||      |||      |||      |||   |||||||||:::
Db        149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||      |||      |||      |||
Db        169 GlyValProGlyGluArgGlyValProGlyAsnThrGlyAlaAlaGlySerAlaGlyAla 188

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||:::||||||      |||   ||||||  ||||||  ||||||
Db        189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy        286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
                ||||||   |||        |||:::|||   |||        ||||||   |||
Db        209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy        337 AAAGCT----------------------------CTGCAAACAGAAATGGCACGTATC 366
              :::                              ||||||        ::::::
Db        229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||       |||      |||:::   ||||||  |||   |||   |||
Db        249 LysLysValGluLeuProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268
```

```
Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
Db        269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
Db        289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
Db        309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
Db        329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySer 348

Qy        658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
Db        349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy        718 CATCTGGCCGTCTGTGAGTTC 738
Db        369 ArgLeuValValCysGluPhe 375
```

RESULT 15
AAW18780
ID AAW18780 standard; protein; 351 AA.
XX
AC AAW18780;
XX
DT 11-NOV-1997 (first entry)
XX
DE Bovine conglutinin.
XX
KW Bovine; conglutinin; recombinant; collagen; neck; truncation;
KW sugar chain; recognition; region; antiviral; neutralisation; virus;
KW treatment; disease; influenza A.
XX
OS Bos taurus.
XX
PN WO9707210-A1.
XX
PD 27-FEB-1997.
XX
PF 25-JAN-1996; 96WO-JP000173.
XX
PR 17-AUG-1995; 95JP-00209698.
PR 02-OCT-1995; 95WO-JP002035.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
DR WPI; 1997-165294/15.
XX
PT Recombinant conglutinin with truncated collagen region - has viral
PT neutralising activity and can be used as an antiviral drug.
XX
PS Disclosure; Page 22-23; 48pp; Japanese.
XX
CC The present sequence is bovine conglutinin (bCG), from which a novel
CC recombinant bCG can be prepared. The recombinant bCG comprises a
CC truncated collagen region of formula Gly-Xaa-Xaa-Gly-Xaa-Xaa (where Xaa
CC is any amino acid), and the neck and sugar chain recognition regions of
CC natural bCG. The recombinant bCG has antiviral activity, i.e. virus
CC neutralising activity, and can be used to treat viral diseases,
CC especially influenza A
XX
SQ Sequence 351 AA;

Alignment Scores:
Pred. No.: 1.32e-32 Length: 351

Score: 442.50 Matches: 91
Percent Similarity: 56.77% Conservative: 39
Best Local Similarity: 39.74% Mismatches: 74
Query Match: 32.42% Indels: 25
DB: 2 Gaps: 6

US-10-054-536-2 (1-747) x AAW18780 (1-351)

```
Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
Db        126 AlaProGlyIleGlnGlyPheProGlyPro---------SerGlyLeuLysGlyGluLys 142

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
Db        143 GlyAlaProGlyGluThrGlyAlaProGlyArgAlaGlyValThrGlyProSerGlyAla 162

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
Db        163 IleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGly 182

Qy        286 GACCCTGGAAAAAGT---CCGGATGGTGATAGTAGCCTGGCTGCC--------------- 327
Db        183 AspProGlyGluThrGlyAlaSerGlyGluSerGlyLeuAlaGluValAsnAlaLeuLys 202

Qy        328 --------------------TCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
Db        203 GlnArgValThrIleLeuAspGlyHisLeuArgArgPheGlnAsnAlaPheSerGlnTyr 222

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
Db        223 LysLysAlaValLeuPheProAspGlyGlnAlaValGlyGluLysIlePheLysThrAla 242

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
Db        243 GlyAlaValLysSerTyrSerAspAlaGluGlnLeuCysArgGluAlaLysGlyGlnLeu 262

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG--------- 537
Db        263 AlaSerProArgSerSerAlaGluAsnGluAlaValThrGlnMetValArgAlaGlnGlu 282

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
Db        283 LysAsnAlaTyrLeuSerMetAsnAspIleSerThrGluGlyArgPheThrTyrProThr 302

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT------GGT 651
Db        303 GlyGluIleLeuValTyrSerAsnTrpAlaAspGlyGluProAsnAsnSerAspGluGly 322

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
Db        323 GlnProGluAsnCysValGluIleProAspGlyLysTrpAsnAspValProCysSer 342

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
Db        343 LysGlnLeuLeuValIleCysGluPhe 351
```

Search completed: June 18, 2004, 21:00:11
Job time : 62.5 secs

Blank

OM nucleic - nucleic search, using sw model

Run on: June 21, 2004, 17:16:09 ; Search time 2543 Seconds
(without alignments)
8771.942 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 747
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	678	90.8	747	29	AY413286	AY413286 Homo sapi
2	671.6	89.9	747	29	AY413287	AY413287 Pan trogl
3	352.2	47.1	735	29	AY413288	AY413288 Mus muscu
4	335	44.8	828	10	BF537636	BF537636 602054516
5	303	40.6	367	9	AV660023	AV660023 AV660023
6	302.2	40.5	657	9	AI255533	AI255533 ui55f12.y
7	301.4	40.3	1162	11	AY325174	AY325174 Rattus no
8	301.4	40.3	1162	11	AY325178	AY325178 Rattus no
c 9	292.4	39.1	869	9	AI195233	AI195233 ui62c03.x
10	291	39.0	367	9	AV660367	AV660367 AV660367
c 11	290.4	38.9	789	9	AI174038	AI174038 ud67a02.x
12	286	38.3	1565	11	AK034788	AK034788 Mus muscu
13	269	36.0	898	11	BC012245	BC012245 Mus muscu
14	268.4	35.9	557	10	BE682267	BE682267 180230 MA
15	268	35.9	788	12	BI147944	BI147944 602912658
c 16	267.6	35.8	784	9	AI173576	AI173576 ud68d03.x
c 17	258.4	34.6	777	9	AI194713	AI194713 ui55f12.x
18	256.6	34.4	664	9	AI876217	AI876217 uj59d03.y
c 19	256	34.3	592	9	AI786738	AI786738 uj56d08.x
20	255.6	34.2	832	14	CB948709	CB948709 AGENCOURT
21	248.4	33.3	627	14	CF179019	CF179019 813103 MA
22	246.2	33.0	738	12	BI144334	BI144334 602908384
23	245.8	32.9	540	13	BX526094	BX526094 BX526094
24	243.4	32.6	753	12	BI145807	BI145807 602909275
25	242.6	32.5	799	9	AI097858	AI097858 ue40g03.y
c 26	239.8	32.1	677	9	AI118440	AI118440 ue40g03.x
27	239.4	32.0	476	10	BF652528	BF652528 276116 MA
28	239	32.0	653	14	CB419261	CB419261 592111 MA
29	236.4	31.6	758	9	AV649316	AV649316 AV649316
30	234.8	31.4	393	9	AV649531	AV649531 AV649531
31	231.6	31.0	760	14	CB947613	CB947613 AGENCOURT
32	228.8	30.6	629	14	CB448619	CB448619 702766 MA
c 33	227.8	30.5	555	14	CB464046	CB464046 725228 MA
c 34	225.6	30.2	696	9	AI876428	AI876428 uj57b07.x
35	218.8	29.3	520	12	BI219712	BI219712 602936130
c 36	216.4	29.0	531	14	CB420319	CB420319 593263 MA
c 37	213.6	28.6	558	9	AI255561	AI255561 ui68e06.x
38	211.6	28.3	588	14	CB462601	CB462601 722700 MA
39	210.6	28.2	815	10	BF232803	BF232803 602023407
c 40	208.8	28.0	558	14	CB448348	CB448348 702478 MA
c 41	207.8	27.8	490	14	CF180387	CF180387 815483 MA
42	206	27.6	700	12	BG084210	BG084210 H3097F11-
43	197	26.4	488	9	AA238657	AA238657 mx82f11.r
44	196.4	26.3	607	9	AI529337	AI529337 ui62c03.y
c 45	195.4	26.2	549	9	AW104610	AW104610 xd84a07.x

ALIGNMENTS

RESULT 1
AY413286
LOCUS AY413286 747 bp DNA linear GSS 17-DEC-2003
DEFINITION Homo sapiens MBL2 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION AY413286
VERSION AY413286.1 GI:39769248
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 747)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 747)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE Direct Submission

```
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1. .>747
                     /gene="MBL2"
                     /locus_tag="HCM4840"
ORIGIN

  Query Match             90.8%;  Score 678;  DB 29;  Length 747;
  Best Local Similarity   90.8%;  Pred. No. 3.3e-180;
  Matches  678;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||
Db        301 CCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
                          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NNNNNNNNNNNNNGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              ||||||||||||||||||||||||||
Db        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747


RESULT 2
AY413287
LOCUS       AY413287                747 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Pan troglodytes MBL2 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY413287
VERSION     AY413287.1  GI:39769249
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 747)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene            <1. .>747
                     /gene="MBL2"
                     /locus_tag="HCM4840"
ORIGIN

  Query Match             89.9%;  Score 671.6;  DB 29;  Length 747;
  Best Local Similarity   90.2%;  Pred. No. 2.1e-178;
  Matches  674;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCTCCTGGAAAGTTGGGGCCTCCAGGA 240

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAGAAAGGAGACCCTGGAAAAAGT 300

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              ||||
Db        301 CCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
                          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NNNNNNNNNNNNNGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||||||||||||||||||||||||||||  |||  |||||||||||||||
Db        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCGTTCGGAATCTCATCAAGGAG 540

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720

Qy        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
              |||||||||||||||||||||||||||
Db        721 CTGGCCGTCTGTGAGTTCCCTATCTGA 747
```

```
RESULT 3
AY413288
LOCUS       AY413288                 735 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Mus musculus MBL2 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY413288
VERSION     AY413288.1  GI:39769250
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 735)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..735
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            <1..>735
                     /gene="MBL2"
                     /locus_tag="HCM4840"
ORIGIN

  Query Match             47.1%;  Score 352.2;  DB 29;  Length 735;
  Best Local Similarity   66.3%;  Pred. No. 4.2e-88;
  Matches  479;  Conservative    0;  Mismatches  237;  Indels    6;  Gaps    2;

Qy         26 TCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCC 85
              ||||||| ||| || ||||| ||| |   |||   |||| ||   | |||    || | ||  |
Db         20 TCCTTCTGCTGTGTGTGGTGACAGTGGTTTATGCAGAGACCTTAACC---GAAGGTGTTC 76

Qy         86 AAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCA 145
              ||||  ||||||||     ||| || |||| || ||||||||| | || ||||||||||||
Db         77 AAAATTCCTGCCCT---GTGGTTACCTGCAGTTCTCCAGGCCTGAATGGCTTCCCAGGCA 133

Qy        146 AAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCT 205
              |||||| ||||| ||  |||||||||||||||| |||||||| |||||||||||||||
Db        134 AAGATGGACGTGACGGTGCCAAGGGAGAAAAGGGAGAACCAGGTCAAGGGCTCAGAGGCT 193

Qy        206 TACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAG 265
               | || ||||| ||||||||   | || ||| |||||      ||||||  | | ||||  ||
Db        194 TGCAAGGCCCTCCTGGAAAAGTAGGACCTACAGGACCCCCAGGGAATCCGGGGTTAAAAG 253

Qy        266 GACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTG 325
              || ||   ||| |   ||||||||||  ||| ||    ||
Db        254 GAGCAGTGGGACCGAAAGGAGACCGTGGGGACAGAGCAGNNNNNNNNNNNNNNNNNNNNN 313

Qy        326 CCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCT 385
                                                            || ||  |||||
Db        314 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGGTGCTCTTCT 373

Qy        386 CTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTG 445
              ||||| |   || ||||||| || ||||   ||    ||| || || | |||   |||| | |||
Db        374 CTCTGAGTGAAAAAGTTGGAAAGAAGTATTTTGTGAGCAGTGTTAAAAAGATGAGCCTTG 433

Qy        446 AAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTG 505
               | | ||||||||| ||||       | | ||||||| ||||||||||||  |||||||||||
Db        434 ACAGAGTGAAGGCCCTGTGCTCCGAATTCCAGGGCTCTGTGGCCACTCCCAGGAATGCTG 493

Qy        506 CAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATG 565
                || ||     ||||| |||||   |   ||| ||     |||| | ||||||||| ||||
Db        494 AGGAAAACTCGGCCATCCAGAAAGTGGCCAAAGATATTGCCTACTTGGGCATCACAGATG 553

Qy        566 AGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGA 625
                || |   |||||    |||| |||||||||||||| ||| ||    ||| || || ||||
Db        554 TGAGGGTTGAAGGCAGTTTTGAGGATCTGACAGGAAACAGAGTGCGCTATACTAATTGGA 613

Qy        626 ACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATG 685
              | || || || ||||||||   | ||      ||  ||||| |||||   || |   || |||||
Db        614 ATGATGGGGAGCCCAACAACACGGGCGATGGGGAAGACTGTGTGGTGATCTTGGGAAATG 673

Qy        686 GCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATCT 745
              || ||||||| || ||||||||||     |||      | ||||   |||||| ||| ||  ||
Db        674 GCAAGTGGAACGATGTCCCCTGCTCTGACTCTTTTTGGCAATCTGTGAATTCTCTGACT 733

Qy        746 GA 747
              ||
Db        734 GA 735
```

```
RESULT 4
BF537636
LOCUS       BF537636                 828 bp    mRNA    linear   EST 11-DEC-2000
DEFINITION  602054516F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4193883 5',
            mRNA sequence.
ACCESSION   BF537636
VERSION     BF537636.1  GI:11625004
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
           Plate: LLAM9526  row: o  column: 04
           High quality sequence stop: 757.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4193883"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_SG2"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             44.8%;  Score 335;  DB 10;  Length 828;
  Best Local Similarity   71.5%;  Pred. No. 3.3e-83;
  Matches  511;  Conservative    0;  Mismatches  195;  Indels    9;  Gaps    5;

Qy         34 CTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACC 93
              ||| || ||||| ||| |  |||  |||| ||  | |||   || | ||  |||||  ||
Db          3 CTGTGTGTGGTGACAGTGGTTTATGCAGAGACCTTAACC---GAAGGTGTTCAAAATTCC 59

Qy         94 TGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGG 153
              ||||||   ||| || |||| || |||||||||| | || ||||||||||||||||||||
Db         60 TGCCCT---GTGGTTACCTGCAGTTCTCCAGGCCTGAATGGCTTCCCAGGCAAAGATGGA 116

Qy        154 CGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGC 213
              ||||| ||  |||||||||||||||| |||||||| |||||||||||||||| || |||
Db        117 CGTGACGGTGCCAAGGGAGAAAAGGGAGAACCAGGTCAAGGGCTCAGAGGCTTGCAAGGC 176

Qy        214 CCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAG 273
              || |||||||| | || ||| |||||  ||||||  | | |||| |  |||| ||  |
Db        177 CCTCCTGGAAAAGTAGGACCTACAGGACCCCCAGGGAATCCGGGGTTAAAAGGAGCAGTG 236

Qy        274 GGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAA 333
              || |  |||||||||| |||  | ||  | ||   ||||| ||||     ||  ||||||
Db        237 GGACCGAAAGGAGACCGTGGGGACAGAGCAGAATTTGATACTAGCGAAATTGATTCAGAA 296

Qy        334 AGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGC 393
              |    ||| || | | ||||  ||  |    | | ||| ||| ||  ||||||||||| |
Db        297 ATTGCAGCCCTACGATCAGAGCTGAGAGCCCTGAGAAACTGGGTGCTCTTCTCTCTGAGT 356

Qy        394 AAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTG 453
              || ||||||| || ||| | ||  ||| || || | |||  |||| | |||| | ||||
Db        357 GAAAAAGTTGGAAAGAAG-TATTTGTGAGCAGTGTTAAAAAGATGAGCCTTGACAGAGTG 415

Qy        454 AAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAAT 513
              |||||| ||||   | | |||||||| |||||||||||| ||||||||||||  || ||
Db        416 AAGGCCCTGTGCTCCGAATTCCAGG-CTCTGTGGCCACTCCCAGGAATGCTGAGGAAAAC 474

Qy        514 GGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACA 573
              |||||  |||||  |  ||| ||   |||| | |||||||||| |||| || |
Db        475 TCGGCCATCCAGAAAGTGGCCAAAGATATTGCCTACTTGGGCATCACAGATGTGAGGGTT 534

Qy        574 GAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGT 633
              |||||    |||| |||||||||||||| ||| ||  ||| || || |||||  || ||
Db        535 GAAGGCAGTTTTGAGGATCTGACAGGAAACAGAGTGCGCTATACTAATTGGAATGATGGG 594

Qy        634 GAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGG 693
              || ||||||||  | ||  ||  |||||  |||||  || |  ||  ||||||| |||||
Db        595 GAGCCCAACAACACGGGCGATGGGGAAGACTGTGTGGTGATCTTGGGAAATGGCAAGTGG 654

Qy        694 AATGACGTCCCCT-GCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATCTGA 747
              || ||  |||||||  ||||   |||  | ||||  |||||||  |||  ||  ||||
Db        655 AACGATGTCCCCTGGCTCTGACTCTTTTTTGGCAATCTGTGAATTCTCTGACTGA 709


RESULT 5
AV660023
LOCUS       AV660023                367 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV660023 GLC Homo sapiens cDNA clone GLCGDC05 3', mRNA sequence.
ACCESSION   AV660023
VERSION     AV660023.1  GI:9881037
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .367
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCGDC05"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             40.6%;  Score 303;  DB 9;  Length 367;
  Best Local Similarity   87.2%;  Pred. No. 2.2e-74;
  Matches  367;  Conservative    0;  Mismatches    0;  Indels   54;  Gaps    1;

Qy        174 AAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCC 233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCC 60

Qy        234 TCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGG 293
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGG 120

Qy        294 AAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGA 353
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGA 180

Qy        354 AATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTT 413
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTT 240

Qy        414 CTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTT 473
              ||||||||||||||||||||||||||||||
Db        241 CTTCCTGACCAATGGTGAAATAATGACCTTTG---------------------------- 272

Qy        474 CCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCAT 533
                                              ||||||||||||||||||||||||||||||
Db        273 ------------------------ATGCTGCAGAGAATGGAGCCATTCAGAATCTCAT 306

Qy        534 CAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCT 593
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCT 366
```

```
Qy        594 G 594
              |
Db        367 G 367
```

RESULT 6
AI255533
LOCUS AI255533 657 bp mRNA linear EST 12-NOV-1998
DEFINITION ui55f12.y1 Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1886351 5' similar to gb:D11440 Mouse mRNA for P28a subunit of Ra-reactive factor, complete (MOUSE);, mRNA sequence.
ACCESSION AI255533
VERSION AI255533.1 GI:3863058
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 657)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:970675
Seq primer: custom primer used
High quality sequence stop: 488.

FEATURES Location/Qualifiers
source 1. .657
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1886351"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse liver mlia"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

ORIGIN

Query Match 40.5%; Score 302.2; DB 9; Length 657;
Best Local Similarity 71.5%; Pred. No. 5.3e-74;
Matches 397; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
Qy        101 CAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATG 160
              | ||| || |||| || ||||||||| | || |||||||||||||||||||| ||||| |
Db         19 CTGTGGTTACCTGCAGTTCTCCAGGCCTGAATGGCTTCCCAGGCAAAGATGGACGTGACG 78

Qy        161 GCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTG 220
              | ||||||||||||||| |||||||| |||||||||||||||||| ||| || |||||||
Db         79 GTGCCAAGGGAGAAAAGGGAGAACCAGGTCAAGGGCTCAGAGGCTTGCAAGGCCCTCCTG 138

Qy        221 GAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAA 280
              |||| | || ||| ||||| |||||| | | |||| | |||| || ||| | |
Db        139 GAAAAGTAGGACCTACAGGACCCCCAGGGAATCCGGGGTTAAAAGGAGCAGTGGGACCGA 198

Qy        281 AAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAG 340
              |||||||| ||| | || | || |||| |||| || ||||||| ||
Db        199 AAGGAGACCGTGGGGACAGAGCAGAATTTGATACTAGCGAAATTGATTCAGAAATTGCAG 258

Qy        341 CTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAG 400
              | || | | |||| || | | ||| ||| || |||||||||| | || |||
Db        259 CCCTACGATCAGAGCTGAGAGCCCTGAGAAACTGGGTGCTCTTCTCTCTGAGTGAAAAAG 318

Qy        401 TTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCT 460
              |||| || |||| || ||| || || | ||| |||| | |||| | ||||||||||
Db        319 TTGGAAAGAAGTATTTTGTGAGCAGTGTTAAAAAGATGAGCCTTGACAGAGTGAAGGCCC 378

Qy        461 TGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCA 520
              |||| | | |||||||| |||||||||||| |||||||||||| || || |||| |
Db        379 TGTGCTCCGAATTCCAGGGCTCTGTGGCCACTCCCAGGAATGCTGAGGAAAACTCGGCCA 438

Qy        521 TTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGC 580
              | ||||| | ||| || |||| | |||||||||| |||| || | |||||
Db        439 TCCAGAAAGTGGCCAAAGATATTGCCTACTTGGGCATCACAGATGTGAGGGTTGAAGGCA 498

Qy        581 AGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCA 640
              |||| |||||||||||||||| ||| || ||| | || ||||| | | ||||||
Db        499 GCTTTGAGGATCTGACAGGAAACAGAGTGCGCTATTCTAATTGGAATGGATGGGAACCCC 558

Qy        641 ACAATGCTGGTTCTG 655
              ||| | || ||
Db        559 ACACCACGGGCGATG 573

RESULT 7
AY325174
LOCUS AY325174 1162 bp mRNA linear HTC 26-JUL-2003
DEFINITION Rattus norvegicus Ab2-011 mRNA, complete cds.
ACCESSION AY325174
VERSION AY325174.1 GI:33086525
KEYWORDS HTC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1162)
AUTHORS Xu,C.S., Li,W.Q., Li,Y.C., Han,H.P., Wang,G.P., Chai,L.Q., Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
TITLE Liver regeneration after PH
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1162)
AUTHORS Xu,C.S., Li,W.Q., Li,Y.C., Han,H.P., Wang,G.P., Chai,L.Q., Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2003) Henan Bioengineering Key Lab, Henan Normal University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R. China

FEATURES Location/Qualifiers
source 1. .1162
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
CDS 41. .1156
/note="liver regeneration-related protein LRRG037"
/codon_start=1
/product="Ab2-011"
/protein_id="AAP92575.1"
/db_xref="GI:33086526"
```

/translation="MSLFTSFLLLCVLTAVYAETLTEGAQSSCPVIACSSPGLNGFPG
KDGHDGAKGEKGEPGQGLRGLQGPPGKVGPAGPPGNPGSKGATGPKGDRGESVEFDTT
NIDLEIAALRSELRAMRKWVLLSMSENVGKKYFMSSVRRMPLNRAKALCSELQGTVAT
PRNAEENRAIQNVAKDVAFLGITDQRTENVFEDLTGNRVRYTNWNEGRLSLRGPELHG
PAANSVYYTPSSSDTENIPPTEETAKDGCALDPYEQSLIHAATESKANSENQHYTLEK
QSVHISVVSVEFLVNLQMFAVTVSSHITRGNRRSLTPGLTHEDFPHKQADSPSLELTG
WLDRMASEPAGCSGLPSCSTRIVDVCQVYLDTMTMFV"

ORIGIN

Query Match 40.3%; Score 301.4; DB 11; Length 1162;
Best Local Similarity 70.6%; Pred. No. 1.3e-73;
Matches 401; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

```
Qy        65 CTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAG 124
             | | |||||     || |       | |   ||    | |||||||||||| || |||||  |
Db        93 CCGAGACCTTAACCGAAGGGGCTCAAAGTAGCTGCCCTGTGATTGCCTGCAGTTCTCCGG 152

Qy       125 GCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAAC 184
             || | ||||||||||||||||||||||| |  || ||  ||||||||||||||| ||||
Db       153 GCCTGAACGGCTTCCCAGGCAAAGATGGACACGACGGTGCCAAGGGAGAAAAGGGAGAAC 212

Qy       185 CAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATC 244
             | || ||||| |||||||||||| |||||||| ||||||||  | || ||| ||||    |
Db       213 CGGGTCAAGGCCTCAGAGGCTTGCAGGGCCCTCCTGGAAAAGTAGGACCTGCAGGGCCCC 272

Qy       245 CAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGG 304
             |||||  | ||||||||  |||| ||| ||| | |||||||||| |||| | |||   |
Db       273 CAGGGAATCCTGGGTCAAAAGGAGCAACGGGACCAAAAGGAGACCGTGGAGAGAGTGTAG 332

Qy       305 ATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTA 364
             |   ||||| || |      ||  | |||||    ||| |||| | | ||  ||  | ||
Db       333 AATTTGATACTACCAACATTGATTTAGAAATTGCAGCCCTGCGATCGGAGCTGAGAGCTA 392

Qy       365 TCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCA 424
             | | ||||||| ||  | | ||| || |  || | ||||| || |||| |||| ||| ||
Db       393 TGAGAAAGTGGGTGCTCCTTTCTATGAGTGAAAATGTTGGAAAGAAGTACTTCATGAGCA 452

Qy       425 ATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTG 484
             || |  ||  ||| || || | | || ||||||  ||||   | |  ||||||  | |||
Db       453 GTGTTAGAAGGATGCCCCTTAACAGAGCGAAGGCTCTGTGCTCCGAACTCCAGGGCACTG 512

Qy       485 TGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAG 544
             |||||||  |||||||||||||  || ||| | ||||| ||||||  |   ||| || |  |
Db       513 TGGCCACTCCCAGGAATGCTGAGGAAAATAGGGCCATCCAGAATGTGGCCAAAGATGTTG 572

Qy       545 CCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATA 604
             |||||  |||||||  || ||  ||| ||| |||       ||||  ||| |||||||||| |
Db       573 CCTTCTTGGGCATAACGGACCAGAGGACTGAAAACGTTTTTGAGGACCTGACAGGAAACA 632

Qy       605 GACTGACCTACACAAACTGGAACGAGGG 632
             || ||  |||||| |||||||| |||||
Db       633 GAGTGCGCTACACTAACTGGAATGAGGG 660
```

RESULT 8
AY325178
LOCUS AY325178 1162 bp mRNA linear HTC 26-JUL-2003
DEFINITION Rattus norvegicus Ab2-001 mRNA, complete cds.
ACCESSION AY325178
VERSION AY325178.1 GI:33086533
KEYWORDS HTC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1162)
AUTHORS Xu,C.S., Li,W.Q., Li,Y.C., Ma,H., Wang,L., Wang,S.F., Han,H.P., Wang,G.P., Chai,L.Q., Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
TITLE Liver regeneration after PH
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1162)
AUTHORS Xu,C.S., Li,W.Q., Li,Y.C., Ma,H., Wang,L., Wang,S.F., Han,H.P., Wang,G.P., Chai,L.Q., Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2003) Henan Bioengineering Key Lab, Henan Normal University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R. China
FEATURES Location/Qualifiers
source 1. .1162
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
CDS 41. .1156
/note="liver regeneration-related protein LRRG053"
/codon_start=1
/product="Ab2-001"
/protein_id="AAP92579.1"
/db_xref="GI:33086534"
/translation="MSLFTSFLLLCVLTAVYAETLTEGAQSSCPVIACSSPGLNGFPG
KDGHDGAKGEKGEPGQGLRGLQGPPGKVGPAGPPGNPGSKGATGPKGDRGESVEFDTT
NIDLEIAALRSELRAMRKWVLLSMSENVGKKYFMSSVRRMPLNRAKALCSELQGTVAT
PRNAEENRAIQNVAKDVAFLGITDQRTENVFEDLTGNRVRYTNWNEGRLSLRGPELHG
PAANSVYYTPSSSDTENIPPTEETAKDGCALDPYEQSLIHAATESKANSENQHYTLEK
QSVHISVVSVEFLVNLQMFAVTVSSHITRGNRRSLTPGLTHEDFPHKQADSPSLELTG
WLDRMASEPAGCSGLPSCSTRIVDVCQVYLDTMTMFV"

ORIGIN

Query Match 40.3%; Score 301.4; DB 11; Length 1162;
Best Local Similarity 70.6%; Pred. No. 1.3e-73;
Matches 401; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

```
Qy        65 CTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAG 124
             | | |||||     || |       | |   ||    | |||||||||||| || |||||  |
Db        93 CCGAGACCTTAACCGAAGGGGCTCAAAGTAGCTGCCCTGTGATTGCCTGCAGTTCTCCGG 152

Qy       125 GCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAAC 184
             || | ||||||||||||||||||||||| |  || ||  ||||||||||||||| ||||
Db       153 GCCTGAACGGCTTCCCAGGCAAAGATGGACACGACGGTGCCAAGGGAGAAAAGGGAGAAC 212

Qy       185 CAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATC 244
             | || ||||| |||||||||||| |||||||| ||||||||  | || ||| ||||    |
Db       213 CGGGTCAAGGCCTCAGAGGCTTGCAGGGCCCTCCTGGAAAAGTAGGACCTGCAGGGCCCC 272

Qy       245 CAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGG 304
             |||||  | ||||||||  |||| ||| ||| | |||||||||| |||| | |||   |
Db       273 CAGGGAATCCTGGGTCAAAAGGAGCAACGGGACCAAAAGGAGACCGTGGAGAGAGTGTAG 332

Qy       305 ATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTA 364
             |   ||||| || |      ||  | |||||    ||| |||| | | ||  ||  | ||
Db       333 AATTTGATACTACCAACATTGATTTAGAAATTGCAGCCCTGCGATCGGAGCTGAGAGCTA 392

Qy       365 TCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCA 424
             | | ||||||| ||  | | ||| || |  || | ||||| || |||| |||| ||| ||
Db       393 TGAGAAAGTGGGTGCTCCTTTCTATGAGTGAAAATGTTGGAAAGAAGTACTTCATGAGCA 452

Qy       425 ATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTG 484
             || |  ||  ||| || || | | || ||||||  ||||   | |  ||||||  | |||
Db       453 GTGTTAGAAGGATGCCCCTTAACAGAGCGAAGGCTCTGTGCTCCGAACTCCAGGGCACTG 512

Qy       485 TGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAG 544
             |||||||  |||||||||||||  || ||| | ||||| ||||||  |   ||| || |  |
Db       513 TGGCCACTCCCAGGAATGCTGAGGAAAATAGGGCCATCCAGAATGTGGCCAAAGATGTTG 572

Qy       545 CCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATA 604
             |||||  |||||||  || ||  ||| ||| |||       ||||  ||| |||||||||| |
Db       573 CCTTCTTGGGCATAACGGACCAGAGGACTGAAAACGTTTTTGAGGACCTGACAGGAAACA 632

Qy       605 GACTGACCTACACAAACTGGAACGAGGG 632
             || ||  |||||| |||||||| |||||
Db       633 GAGTGCGCTACACTAACTGGAATGAGGG 660
```

RESULT 9

```
AI195233/c
LOCUS       AI195233                869 bp    mRNA    linear   EST 14-OCT-1998
DEFINITION  ui62c03.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1886980 3' similar to gb:X15422_cds1 MANNOSE-BINDING PROTEIN
            C PRECURSOR (HUMAN); gb:D11440 Mouse mRNA for P28a subunit of
            Ra-reactive factor, complete (MOUSE);, mRNA sequence.
ACCESSION   AI195233
VERSION     AI195233.1  GI:3747839
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:971304
            Seq primer: custom primer used
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1886980"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse liver mlia"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
ORIGIN
```

Query Match 39.1%; Score 292.4; DB 9; Length 869;
Best Local Similarity 67.0%; Pred. No. 3.7e-71;
Matches 424; Conservative 0; Mismatches 208; Indels 1; Gaps 1;

```
Qy        115 AGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAA 174
              || |||| |||| | || | || || ||  ||     || ||| | ||  | |||||||||
Db        805 AGTTCTCAAGGCCTGAATGCCTACCNAGCAAAGATGGGACGTNACGGTGCAAAGGGAGAA 746

Qy        175 AAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCT 234
              ||||  ||| |||| ||||||||||||  |   |     | ||| |||  |||    || |
Db        745 AAGGAAGAAGCAGGTCAAGGGCTCAGATGGCTTGCAAGGCCCTCTGNNAAGAGTAGGACG 686

Qy        235 CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA 294
                  | |  |   ||   || |||| |  |||| ||  |  || ||||||||||| ||
```

```
Db        685 TACAGGACCCNCAGGGAATCGGGGTTAAAAGGAGCAGTGTACCGAAAAGGAGACCGTGNG 626

Qy        295 AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
              | ||  | ||   ||||| ||||     ||  |||||||    ||| || | | ||||
Db        625 GACAGAGCAGAATTTGATACTAGCGAAATTGATTCAGAAATTGCAGCCCTACGATCAGAG 566

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
              ||  |    | | |||||||| ||  |||||||||| |  |  ||| ||| || ||||
Db        565 CTGAGAGCCCTGAGAAAGTGGGTGCTCTTCTCTCTGAGTGANAAAG-TGGAAAGAAGTAT 507

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              ||  ||| || || | |||  |||| | |||| | |||||||||| ||||   | | |||
Db        506 TTTGTGAGCAGTGTTAAAAAGATGAGCCTTGACAGAGTGAAGGCCCTGTGCTCCGAATTC 447

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
              |||| ||||||||||| |||||||||||| || ||    ||||| ||||| |   |
Db        446 CAGGGCTCTGTGGCCACTCCCAGGAATGCTGAGGAAAACTCGGCCATCCAGAAAGTGGCC 387

Qy        535 AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTG 594
              || ||    |||| | |||||||||| |||| || |   |||||    |||| |||||||
Db        386 AAAGATATTGCCTACTTGGGCATCACAGATGTGAGGGTTGAAGGCAGTTTTGAGGATCTG 327

Qy        595 ACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCT 654
              |||||||| ||| ||  ||| || || ||||| || || || ||||||||  | ||   |
Db        326 ACAGGAAACAGAGTGCGCTATACTAATTGGAATGATGGGGAGCCCAACAACACGGGCGAT 267

Qy        655 GATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACC 714
              |  ||||| |||||  || |  ||  ||||||| ||||||| || ||||||||||   |
Db        266 GGGGAAGACTGTGTGGTGATCTTGGGAAATGGCAAGTGGAACGATGTCCCCTGCTCTGAC 207

Qy        715 TCCCATCTGGCCGTCTGTGAGTTCCCTATCTGA 747
              ||   | ||||  ||||||| ||| ||  ||||
Db        206 TCTTTTTTGGCAATCTGTGAATTCTCTGACTGA 174
```

RESULT 10
AV660367

```
LOCUS       AV660367                367 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV660367 GLC Homo sapiens cDNA clone GLCGGH07 3', mRNA sequence.
ACCESSION   AV660367
VERSION     AV660367.1  GI:9881381
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .367
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCGGH07"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
```

/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 39.0%; Score 291; DB 9; Length 367;
Best Local Similarity 86.9%; Pred. No. 5.4e-71;
Matches 366; Conservative 0; Mismatches 0; Indels 55; Gaps 2;

```
Qy      174 AAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCC 233
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 AAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCC 60

Qy      234 TCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGG 293
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 TCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGG 120

Qy      294 AAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGA 353
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 AAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGA 180

Qy      354 AATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTT 413
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTT 240

Qy      414 CTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTT 473
            ||||||||||||||||||||||||||||||||
Db      241 CTTCCTGACCAATGGTGAAATAATGACCTTTG---------------------------- 272

Qy      474 CCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCAT 533
                                    ||||||||||||||||||||||||||||||||||
Db      273 ------------------------ATGCTGCAGAGAATGGAGCCATTCAGAATCTCAT 306

Qy      534 CAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCT 593
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      307 CAAGGAGGAAGCCTTCCT-GGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCT 365

Qy      594 G 594
            |
Db      366 G 366
```

RESULT 11
AI174038/c
LOCUS AI174038 789 bp mRNA linear EST 07-OCT-1998
DEFINITION ud67a02.x1 Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1450922 3' similar to gb:D11440 Mouse mRNA for P28a subunit of Ra-reactive factor, complete (MOUSE);, mRNA sequence.
ACCESSION AI174038
VERSION AI174038.1 GI:3720180
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 789)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:924238
Seq primer: custom primer used
High quality sequence stop: 486.

FEATURES Location/Qualifiers
source 1. .789
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1450922"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse liver mlia"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

ORIGIN

Query Match 38.9%; Score 290.4; DB 9; Length 789;
Best Local Similarity 69.8%; Pred. No. 1.3e-70;
Matches 432; Conservative 0; Mismatches 184; Indels 3; Gaps 3;

```
Qy      130 AACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGC 189
            || ||||| |||| |||||||  |||  ||  |||||||||||||| | ||||||  |
Db      789 AATGGCTTTCCAGCAAAAGATGACCGT-NCGGTGCCAAGGGAGAAAAAGNAGAACCA-GT 732

Qy      190 CAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGG 249
            ||||||  | ||||||  || |||||  ||||||||  | ||  |||  |||||   ||||||
Db      731 CAAGGGTCAANAGGCTTGCAAGGCCCTCCTGGAAAAGTAGGACCTACAGGACCCCCAGGG 672

Qy      250 CCTTCTGG-GTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGG 308
              | | ||  ||  |  ||||  ||  |||  |  ||||||||||  |||  |  ||  | ||
Db      671 GATCCGGGNGTTAAAAGGAGCAGTGGGACCGAAAGGAGACCGTGGGGACAGAGCAGAATT 612

Qy      309 TGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAA 368
            |||||  |||     ||  |||||||   |||  ||  |  |  ||||   ||   |  | ||
Db      611 TGATACTAGCGAAATTGATTCAGAAATTGCAGCCCTACGATCAGAGCTGAGAGCCCTGAG 552

Qy      369 AAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGG 428
            |||  |||  ||  ||||||||||  |  ||  |||||||  ||  ||||  ||  |||  ||  ||
Db      551 AAACTGGGTGCTCTTCTCTCTGAGTGAAAAAGTTGGAAAGAAGTATTTTGTGAGCAGTGT 492

Qy      429 TGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGC 488
            |  |||  ||||  |  ||||  |  ||||||||||  ||||   |  |  |||||||  ||||||||||
Db      491 TAAAAAGATGAGCCTTGACAGAGTGAAGGCCCTGTGCTCCGAATTCCAGGGCTCTGTGGC 432

Qy      489 CACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTT 548
            |||  ||||||||||||||   ||  ||     |||||  |||||   |   |||  ||    ||||
Db      431 CACTCCCAGGAATGCTGAGGAAAACTCGGCCATCCAGAAAGTGGCCAAAGATATTGCCTA 372

Qy      549 CCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACT 608
            |  ||||||||||  ||||  ||  |   |||||     ||||  |||||||||||||||  |||  |
Db      371 CTTGGGCATCACAGATGTGAGGGTTGAAGGCAGTTTTGAGGATCTGACAGGAAACAGAGT 312

Qy      609 GACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGT 668
            |  |||  ||  ||  |||||  ||  ||  ||  ||||||||  |  ||   ||   |||||  |||||
Db      311 GCGCTATACTAATTGGAATGATGGGGAGCCCAACAACACGGGCGATGGGGAAGACTGTGT 252

Qy      669 ATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGT 728
              ||  |  ||  |||||||  |||||||  ||  ||||||||||   |||   |  ||||  |
Db      251 GGTGATCTTGGGAAATGGCAAGTGGAACGATGTCCCCTGCTCTGACTCTTTTTTGGCAAT 192
```

```
Qy        729 CTGTGAGTTCCCTATCTGA 747
              ||||||| ||| ||  ||||
Db        191 CTGTGAATTCTCTGACTGA 173
```

RESULT 12
AK034788
LOCUS AK034788 1565 bp mRNA linear HTC 18-SEP-2003
DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430039D04 product:mannose binding lectin, liver (A), full insert sequence.
ACCESSION AK034788
VERSION AK034788.1 GI:26084196
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1565)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1. .1565
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9430039D04"
/db_xref="MGI:2398998"
/db_xref="taxon:10090"
/clone="9430039D04"
/tissue_type="embryonic body between diaphragm region and neck"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
misc_feature 120. .840
/note="mannose binding lectin, liver (A) (MGD|MGI:96923, GB|NM_010775, evidence: BLASTN, 99%, match=922) putative"
ORIGIN

Query Match 38.3%; Score 286; DB 11; Length 1565;
Best Local Similarity 68.0%; Pred. No. 3.5e-69;
Matches 416; Conservative 0; Mismatches 190; Indels 6; Gaps 1;

```
Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGA 201
              |||| ||||||| | |||||  |||| || || ||||| |||||||| ||||||||||
Db        229 GGCAGAGATGGGAGAGATGGACCCAAAGGGGAGAAGGGAGAACCAGGTCAAGGGCTCAGG 288

Qy        202 GGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCA 261
              ||||| |||||||| || || || |||||||||||||||| |   ||    ||||
Db        289 GGCTTGCAGGGCCCTCCAGGGAAATTGGGGCCTCCAGGAAGTGTTGGAAGCCCTGGAAGT 348

Qy        262 CCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTG 321
              |||||||||||| |||||||| || |||| |||| | | |   |   |||   |  |||
Db        349 CCAGGACCAAAAGGCCAAAAGGGGGACCATGGAGACAATAGAGCCATTGAGGAGAAGCTG 408

Qy        322 GC------TGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGG 375
              ||      |  |||| | ||     ||| || || || ||   |  | ||| |||| |
Db        409 GCAAATATGGAGGCAGAGATAAGGATCCTGAAATCAAAACTGCAGCTAACCAACAAGTTG 468

Qy        376 CTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATA 435
              |  ||||||| |||||||| |  |||||| ||||| ||  |||||||  ||| |
Db        469 CATGCCTTCTCAATGGGCAAAAAGTCTGGGAAGAAGTTGTTTGTGACCAACCATGAGAAG 528

Qy        436 ATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCC 495
              ||| |||||  |||||||| | ||||    || |||| | | ||||||| | ||||
Db        529 ATGCCCTTTTCCAAAGTGAAGTCTCTGTGCACAGAGCTCCAAGGCACTGTGGCTATCCCC 588

Qy        496 AGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGC 555
              |||||||||| ||||||    |||||||| | |  ||  |    |||||||| |||
Db        589 AGGAATGCTGAAGAGAACAAGGCCATTCAAGAAGTGGCCACAGGCATTGCCTTCCTAGGC 648

Qy        556 ATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTAC 615
              ||||| || |||  ||| |||||||||| || |  |||||||    || || ||||||
Db        649 ATCACGGACGAGGCGACTGAAGGGCAGTTCATGTACGTGACAGGGGGGAGGCTCACCTAC 708

Qy        616 ACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTA 675
              |  |||||||  ||| ||| || || ||   ||| |||| |||||  |||||  |  ||
Db        709 AGCAACTGGAAAAAGGATGAGCCAAATAACCATGGCTCTGGGGAAGACTGTGTCATTATA 768

Qy        676 CTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
              |  | |||||  |||||||||| | ||||    | |||    ||| ||||| |||
```

```
Db        769 TTAGATAATGGTTTGTGGAATGACATTTCCTGTCAAGCTTCCTTCAAGGCTGTCTGCGAG 828

Qy        736 TTCCCTATCTGA 747
              ||||||  ||||
Db        829 TTCCCAGCCTGA 840


RESULT 13
BC012245
LOCUS       BC012245                898 bp    mRNA    linear   HTC 20-SEP-2002
DEFINITION  Mus musculus, Similar to mannose binding lectin, liver (A), clone
            IMAGE:4193682, mRNA.
ACCESSION   BC012245
VERSION     BC012245.1  GI:15126625
KEYWORDS    HTC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 24 Row: j Column: 9
            This clone has the following problem: frame shifted.
FEATURES              Location/Qualifiers
     source           1. .898
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:4193682"
                      /tissue_type="Salivary gland, 10 week old female mouse"
                      /clone_lib="NCI_CGAP_SG2"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
ORIGIN

  Query Match             36.0%;  Score 269;  DB 11;  Length 898;
  Best Local Similarity   67.2%;  Pred. No. 1.6e-64;
  Matches  416;  Conservative    0;  Mismatches  190;  Indels   13;  Gaps    2;

Qy        142 GGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGG-------CCAAGG 194
              |||| ||||||| | |||||  |||| || || ||||| ||||||||       |||||
Db        184 GGCAGAGATGGGAGAGATGGACCCAAAGGGGAGAAGGGAGAACCAGGTATAGAGTCAAGG 243

Qy        195 GCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTC 254
              |||||| ||||| |||||||| || || || ||||||||||||||| |    ||     |
Db        244 GCTCAGGGGCTTGCAGGGCCCTCCAGGGAAATTGGGGCCTCCAGGAAGTGTTGGAAGCCC 303

Qy        255 TGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAG 314
              |||    |||||||||| |||||||| || |||| |||| | | |    |    |||
Db        304 TGGAAGTCCAGGACCAAAAGGCCAAAAGGGGGACCATGGAGACAATAGAGCCATTGAGGA 363

Qy        315 TAGCCTGGC------TGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAA 368
              | |||||       |   |||| | ||     ||| || || || ||   |  | |||
Db        364 GAAGCTGGCAAATATGGAGGCAGAGATAAGGATCCTGAAATCAAAACTGCAGCTAACCAA 423

Qy        369 AAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGG 428
              |||| ||   ||||||| |||||||| |   |||||| ||||| ||  |||||||
Db        424 CAAGTTGCATGCCTTCTCAATGGGCAAAAAGTCTGGGAAGAAGTTGTTTGTGACCAACCA 483

Qy        429 TGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGC 488
              ||| |  ||| |||||  |||||||||  |  ||||     ||  |||| |  | |||||||
Db        484 TGAGAAGATGCCCTTTTCCAAAGTGAAGTCTCTGTGCACAGAGCTCCAAGGCACTGTGGC 543

Qy        489 CACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTT 548
              | ||||||||||||||| ||||||   ||||||||  |  |   ||  |      |||||
Db        544 TATCCCCAGGAATGCTGAAGAGAACAAGGCCATTCAAGAAGTGGCCACAGGCATTGCCTT 603

Qy        549 CCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACT 608
              ||| |||||||| || |||  ||| ||||||||||||  || |  |||||||    || ||
Db        604 CCTAGGCATCACGGACGAGGCGACTGAAGGGCAGTTCATGTACGTGACAGGGGGGAGGCT 663

Qy        609 GACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGT 668
              |||||||  |||||||| ||| ||| || || ||   ||| |||| ||||| |||||
Db        664 CACCTACAGCAACTGGAAAAAGGATGAGCCAAATAACCATGGCTCTGGGGAAGACTGTGT 723

Qy        669 ATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGT 728
              |  || |  | |||||  |||||||||| |  ||||     | |||     ||| ||
Db        724 CATTATATTAGATAATGGTTTGTGGAATGACATTTCCTGTCAAGCTTCCTTCAAGGCTGT 783

Qy        729 CTGTGAGTTCCCTATCTGA 747
              ||| ||||||||   ||||
Db        784 CTGCGAGTTCCCAGCCTGA 802


RESULT 14
BE682267
LOCUS       BE682267                557 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  180230 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE682267
VERSION     BE682267.1  GI:10067954
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,
            Quackenbush,J. and Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 83  row: O  column: 15
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .557
```

```
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
ORIGIN

  Query Match             35.9%;  Score 268.4;  DB 10;  Length 557;
  Best Local Similarity   68.9%;  Pred. No. 1.7e-64;
  Matches  384;  Conservative    0;  Mismatches  167;  Indels    6;  Gaps    1;

Qy        152 GGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGG 211
              |||| |||||  |||||||||||||||| || ||||| ||||||||||||||||||  ||||
Db          1 GGCGAGATGGACCCAAGGGAGAAAAGGGAGAGCCAGGGCAAGGGCTCAGAGGCTCGCAGG 60

Qy        212 GCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAA 271
              |||| || || ||  ||||||||||||||||    ||||  | |||||   |||||||| |
Db         61 GCCCTCCAGGGAAAATGGGGCCTCCAGGAAACATAGGGAATCCTGGGCTTCCAGGACCCA 120

Qy        272 AGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAG 331
              ||||  | ||||||||  | ||||  | ||  ||| || |||    ||  ||||||  ||  |
Db        121 GGGGCTACAAAGGAGATCGTGGAGATAGCTCGGTTGCTGAGGCTAAGCTGGCTAGCTTGG 180

Qy        332 AAAG------AAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCT 385
              | ||      ||  |  |||| | ||||| |||  |   |||||||||  ||  |||||||
Db        181 AGAGACAGATAAGGGACCTGCGATCAGAACTGGATCACGTCAAAAAGTTGCAAACCTTCT 240

Qy        386 CTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTG 445
              |  |||||||| |  ||||||  ||| | |   ||||||||  ||||||  |||  | |||
Db        241 CCTTGGGCAAAAAGTCTGGGAAGAAGCTGTATGTGACCAATCGTGAAAAGATGCCTTTTT 300

Qy        446 AAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTG 505
              |  ||||||||  ||||       |   |||  | ||||||||  ||||  ||||||  |
Db        301 CCAGCGTGAAGGCTCTGTGCACTGCACTTGGGGCTACCGTGGCCACACCCAAGAATGCAG 360

Qy        506 CAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATG 565
              |||||   ||||||  |||  |  |  |   ||  ||||||||||||||||  ||||
Db        361 AGGAGAACAAAGCCATCCAGGACATGGCCTCTGATACCGCCTTCCTGGGCATCACAGATG 420

Qy        566 AGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGA 625
              ||  |||  ||||||||||||  ||  ||  |  ||  |||    ||  ||   ||||||  |||||||
Db        421 AGGTGACTGAAGGGCAGTTTATGTATGTAACTGGAGGAAGGCTAGGCTACAGCAACTGGA 480

Qy        626 ACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATG 685
              |  ||  |||||||||  ||   |||  ||    ||  ||  |||||    ||  |   | | |
Db        481 AGAAGAATGAACCCAATAACTATGGCTCANGGGAGGACTGTGTGAGCCTCTTACCAGACG 540

Qy        686 GCCAGTGGAATGACGTC 702
              | |  |||||||||  ||
Db        541 GGCTCTGGAATGACATC 557


RESULT 15
BI147944
LOCUS       BI147944                788 bp    mRNA    linear   EST 05-JUL-2001
DEFINITION  602912658F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5053406 5',
            mRNA sequence.
ACCESSION   BI147944
VERSION     BI147944.1  GI:14607945
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11145  row: d  column: 15
            High quality sequence stop: 788.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5053406"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             35.9%;  Score 268;  DB 12;  Length 788;
  Best Local Similarity   72.4%;  Pred. No. 2.8e-64;
  Matches  417;  Conservative    0;  Mismatches  150;  Indels    9;  Gaps    5;

Qy         26 TCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCC 85
              ||||||| ||| || ||||| ||| |  |||  |||| ||  | |||   || | ||  |
Db        205 TCCTTCTGCTGTGTGTGGTGACAGTGGTTTATGCAGAGACCTTAACC---GAAGGTGTTC 261

Qy         86 AAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCA 145
              ||||  ||||||||    ||| || |||| || |||||||||  | ||  ||||||||||||||
Db        262 AAAATTCCTGCCCT---GTGGTTACCTGCAGTTCTCCAGGCCTGAATGGCTTCCCAGGCA 318

Qy        146 AAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCT 205
              |||||||  |||||  ||  |||||||||||||||  ||||||||  ||||||||||||||||
Db        319 AAGATGGACGTGACGGTGCCAAGGGAGAAAAGGGAGAACCAGGTCAAGGGCTCAGAGGCT 378

Qy        206 TACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAG 265
              |  ||  |||||  ||||||||   |  ||  |||  |||||   ||||||  |  |  ||||  |  ||
Db        379 TGCAAGGCCCTCCTGGAAAAGTAGGACCTACAGGACCCCCAGGGAATCCGGGGTTAAAAG 438

Qy        266 GACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTG 325
              ||  ||  |||  |  ||||||||||  |||   |  ||   |  ||    |||||  ||||      ||
Db        439 GAGCAGTGGGACCGAAAGGAGACCGTGGGGACAGAGCAGAATTTGATACTAGCGAAATTG 498

Qy        326 CCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCT 385
              |||||||     |||  ||  |  |  ||||   ||   |     |  |  |||  |||  ||   |||||
Db        499 ATTCAGAAATTGCAGCCCTACGATCAGAGCTGAGAGCCCTGAGAAACTGGGTGCTCTTCT 558

Qy        386 CTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTG 445
              |||||  |   ||  |||||||  ||  |||  |  ||   |||  ||  ||  |  |||   ||||  |  |||
Db        559 CTCTGAGTGAAAAAGTTGGAAAGAAG-TATTTGTGAGCAGTGTTAAAAAGATGAGCCTTG 617

Qy        446 AAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTG 505
              |  |  ||||||||||  ||||    |  |  ||||||||  ||||||||||  ||||||||||||
Db        618 ACAGAGTGAAGGCCCTGTGCTCCGAATTCCAGGGCTCTGTGGCCACTCCCAGGAATGCTG 677

Qy        506 CAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATG 565
              ||  ||     |  |||  |||||   |   |||  ||     ||||  |  ||||||||||  ||||
Db        678 -AGGAAAACTCGGCATCCAGAAAGTGGCCAAAGATATTGCCTACTTGGGCATCACAGATG 736

Qy        566 AGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAA 601
              ||  |    |||  |||||||||  ||||||||||||||
Db        737 TGAGGGTTGAA-GGCAGTTTGAGGATCTGACAGGAA 771


Search completed: June 21, 2004, 19:06:08
Job time : 2549 secs
```

Blank

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 18, 2004, 20:56:36 ; Search time 49.5 Seconds
(without alignments)
9522.913 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 1365
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10054536/runat_18062004_173455_13584/app_query.fasta_1.903
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054536_@CGN_1_1_86_@runat_18062004_173455_13584 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1073	78.6	236	6	Q28518	Q28518 macaca mula
2	814.5	59.7	245	6	Q28517	Q28517 macaca mula
3	808	59.2	240	6	Q9XSW3	Q9xsw3 sus scrofa
4	684	50.1	224	11	Q9Z294	Q9z294 rattus sp.
5	636	46.6	371	11	Q7TMC0	Q7tmc0 rattus norv
6	551.5	40.4	254	13	Q98TA4	Q98ta4 gallus gall
7	540.5	39.6	238	13	O57451	O57451 gallus gall
8	454	33.3	375	4	Q86YK9	Q86yk9 homo sapien
9	450	33.0	375	4	Q8TCD8	Q8tcd8 homo sapien
10	428	31.4	246	13	Q9I9Q7	Q9i9q7 carassius a
11	422.5	31.0	369	6	Q863A1	Q863a1 bos taurus
12	419.5	30.7	248	6	Q9TT06	Q9tt06 ovis aries
13	418.5	30.7	256	13	Q9I9Q9	Q9i9q9 cyprinus ca
14	411	30.1	378	6	Q9N1X4	Q9n1x4 sus scrofa
15	407	29.8	248	6	Q95L88	Q95l88 equus cabal
16	407	29.8	248	6	Q9N0G1	Q9n0g1 equus cabal
17	406.5	29.8	251	13	Q9I9Q8	Q9i9q8 brachydanio
18	396	29.0	248	11	Q9CQI1	Q9cqi1 mus musculu
19	394.5	28.9	248	4	Q8IWL2	Q8iwl2 homo sapien
20	388.5	28.5	248	4	Q8IWL1	Q8iwl1 homo sapien
21	326	23.9	62	4	Q9HCS8	Q9hcs8 homo sapien
22	304.5	22.3	271	4	Q9BWP8	Q9bwp8 homo sapien
23	298.5	21.9	272	11	Q9DC75	Q9dc75 mus musculu
24	295.5	21.6	271	13	Q7T0T0	Q7t0t0 xenopus lae
25	295	21.6	742	11	Q8K4Q8	Q8k4q8 mus musculu
26	295	21.6	742	11	Q8C979	Q8c979 mus musculu
27	289.5	21.2	277	4	Q9Y6Z7	Q9y6z7 homo sapien
28	289	21.2	742	11	Q8VIF6	Q8vif6 mus musculu
29	280.5	20.5	277	11	Q8CF98	Q8cf98 mus musculu
30	280.5	20.5	417	4	Q8TCR2	Q8tcr2 homo sapien
31	278.5	20.4	277	11	Q8C1C5	Q8c1c5 mus musculu
32	278.5	20.4	742	4	Q9BYH7	Q9byh7 homo sapien
33	278.5	20.4	742	4	Q8WZA4	Q8wza4 homo sapien
34	275	20.1	268	4	Q7Z6N1	Q7z6n1 homo sapien
35	248	18.2	116	6	Q9N1X3	Q9n1x3 sus scrofa
36	239.5	17.5	88	6	Q8MJA5	Q8mja5 macaca mula
37	237.5	17.4	165	6	Q9TUC5	Q9tuc5 macaca mula
38	234	17.1	117	4	Q8TC19	Q8tc19 homo sapien
39	219	16.0	222	13	Q90XB2	Q90xb2 gallus gall
40	202	14.8	309	11	Q8VH32	Q8vh32 mus musculu
41	201.5	14.8	309	11	Q8VH33	Q8vh33 mus musculu
42	199	14.6	104	6	Q9XSP4	Q9xsp4 sus scrofa
43	198.5	14.5	101	6	Q9XSQ2	Q9xsq2 ovis aries
44	198	14.5	271	5	Q9VET6	Q9vet6 drosophila
45	198	14.5	322	5	Q86NZ7	Q86nz7 drosophila

ALIGNMENTS

RESULT 1
Q28518
ID Q28518 PRELIMINARY; PRT; 236 AA.
AC Q28518;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mannose-binding protein C (Fragment).
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97031450; PubMed=8877375;
RA Mogues T., Ota T., Tauber A.I., Sastry K.N.;
RT "Characterization of two mannose-binding protein cDNAs from rhesus
RT monkey (Macaca mulatta): structure and evolutionary implications.";
RL Glycobiology 6:543-550(1996).
DR EMBL; L43911; AAB48071.1; -.
DR HSSP; P11226; 1HUP.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR008160; Collagen.

```
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   236 AA;  24911 MW;  449CBE887C89732E CRC64;

Alignment Scores:
Pred. No.:                4.63e-93      Length:          236
Score:                    1073.00       Matches:         195
Percent Similarity:       97.22%        Conservative:    15
Best Local Similarity:    90.28%        Mismatches:      6
Query Match:              78.61%        Indels:          0
DB:                       6             Gaps:            0

US-10-054-536-2 (1-747) x Q28518 (1-236)

Qy         97 CCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGT 156
              |||   |||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         21 ProGluValIleAlaCysAsnSerProGlyIleAsnGlyPheProGlyLysAspGlyArg 40

Qy        157 GATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCC 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AspGlyThrLysGlyGluLysGlyGluProGlyGlnGlyLeuArgGlyLeuGlnGlyPro 60

Qy        217 CCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGC 276
              |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db         61 ProGlyLysLeuGlyProProGlyAsnProGlySerSerGlySerProGlyProLysGly 80

Qy        277 CAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGA 336
              ||||||||||||||||||:::|||||||||   :::|||||||||||||||||||||||
Db         81 GlnLysGlyAspProGlyGluSerProAspCysGluSerSerLeuAlaAlaSerGluArg 100

Qy        337 AAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAA 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        101 LysAlaLeuGlnThrGluMetAlaArgIleLysLysTrpLeuThrPheSerLeuGlyArg 120

Qy        397 CAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAG 456
              |||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db        121 GlnValGlyAsnLysPhePheLeuThrAsnGlyGluMetMetThrPheAspLysValLys 140

Qy        457 GCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGA 516
              |||||||||   :::|||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaLeuCysAlaArgPheGlnAlaSerValAlaThrProArgAsnAlaAlaGluAsnArg 160

Qy        517 GCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAA 576
              |||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        161 AlaIleGlnAsnLeuIleLysGluGluAlaPheLeuGlyIleThrAspGluAsnThrGlu 180

Qy        577 GGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAA 636
              |||:::|||||||||||||||||||:::|||||||||||||||||||||:::||||||
Db        181 GlyGluPheValAspLeuThrGlyAsnLysLeuThrTyrThrAsnTrpAsnAspGlyGlu 200

Qy        637 CCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAAT 696
              |||||||||||||||||:::|||||||||||||||||||||||||||||:::||||||
Db        201 ProAsnAsnAlaGlySerAsnGluAspCysValLeuLeuLeuLysAsnGlyLysTrpAsn 220

Qy        697 GACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATC 744
              |||:::|||||||||:::|||||||||||||:::|||||||||||||||
Db        221 AspIleProCysSerSerSerHisLeuAlaLeuCysGluPheProIle 236

RESULT 2
Q28517
ID   Q28517      PRELIMINARY;      PRT;   245 AA.
AC   Q28517;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mannose-binding protein A precursor (Fragment).
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9544;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97031450; PubMed=8877375;
RA   Mogues T., Ota T., Tauber A.I., Sastry K.N.;
RT   "Characterization of two mannose-binding protein cDNAs from rhesus
RT   monkey (Macaca mulatta): structure and evolutionary implications.";
RL   Glycobiology 6:543-550(1996).
DR   EMBL; L43912; AAB48070.1; -.
DR   HSSP; P19999; 1YTT.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Signal.
FT   NON_TER       1       1
FT   SIGNAL       <1      16       POTENTIAL.
FT   CHAIN        17     245       MANNOSE-BINDING PROTEIN A.
SQ   SEQUENCE   245 AA;  25693 MW;  AFAF0FCEF3DE6146 CRC64;

Alignment Scores:
Pred. No.:                1.62e-68      Length:          245
Score:                    814.50        Matches:         155
Percent Similarity:       75.92%        Conservative:    31
Best Local Similarity:    63.27%        Mismatches:      56
Query Match:              59.67%        Indels:          3
DB:                       6             Gaps:            2

US-10-054-536-2 (1-747) x Q28517 (1-245)

Qy         13 CCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACC 72
              ||||||   |||:::|||||||||||:::|||   ||||||   |||||||||
Db          1 ProSerPheProValLeuLeuLeuSerValValThrAlaSerCysSerGluThrLysAla 20

Qy         73 TGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAAC 132
              |||||||||||||||||||||||   :::||||||   |||      |||      |||
Db         21 CysGluAspAlaGlnLysThrCys---SerValIleThrCysGlyIleProValThrAsn 39

Qy        133 GGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAA 192
              |||   ||||||:::|||||||||||||   |||||||||||||||||||||||||||||
Db         40 GlyThrProGlyArgAspGlyArgAspGlyProLysGlyGluLysGlyGluProGlyGln 59

Qy        193 GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCT 252
              |||||||||||||||||||||||||||||||   |||||||||||||||   |||
Db         60 GlyLeuArgGlyLeuGlnGlyProProGlyLysSerGlyProProGlyAsnThrGlyAla 79

Qy        253 TCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGAT 312
                 |||   |||||||||:::|||||||||||||||   |||   :::      :::
Db         80 ProGlyIleProGlyProArgGlyGlnLysGlyAspHisGlyAspAsnSerValAlaGlu 99

Qy        313 AGTAGCCTGGCTGCCTCAGAAAGA------AAAGCTCTGCAAACAGAAATGGCACGTATC 366
              :::   ||||||      ||||||      :::::|||:::::|||:::      :::
Db        100 AlaLysLeuAlaAsnLeuGluArgGlnLeuGlnSerLeuArgSerGluLeuAspHisMet 119

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||      ||||||||||||||      |||   |||   |||:::||||||
Db        120 LysLysLeuGlnAlaPheSerLeuGlyLysMetSerGlyLysLysLeuPheValThrAsn 139

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              ||||||   |||   |||   :::|||||||||||||||      ||||||:::|||
Db        140 GlyGluArgMetProPheSerGluValLysAlaLeuCysAlaGlyLeuGlnAlaThrVal 159

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAAGCC 546
              |||   |||:::||||||   ||||||   |||||||||:::::   |||:::   |||
Db        160 AlaAlaProLysAsnAlaGluGluAsnLysAlaIleGlnAspValAlaLysAspThrAla 179
```

```
Qy        547 TTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGA 606
              ||||||||||||||||||||  ||||||||||||||:::  :::||||||  |||
Db        180 PheLeuGlyIleThrAspGluAlaThrGluGlyGlnPheMetTyrValThrGlyGlyArg 199

Qy        607 CTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGT 666
              |||||||||:::||||||  :::  |||||||||:::  ||||||  |||||||||
Db        200 LeuThrTyrSerAsnTrpLysLysAspGluProAsnAspHisGlySerGlyGluAspCys 219

Qy        667 GTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCC 726
              |||:::||||||  ||||||  |||||||||:::  |||:::  |||::::::|||
Db        220 ValIleLeuLeuSerAsnGlyLeuTrpAsnAspIleSerCysThrAlaSerTyrIleAla 239

Qy        727 GTCTGTGAGTTCCCT 741
              |||||||||||||||
Db        240 ValCysGluPhePro 244
```

RESULT 3
Q9XSW3
ID Q9XSW3 PRELIMINARY; PRT; 240 AA.
AC Q9XSW3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mannose-binding lectin.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Agah A., Young K., Stahl G.L.;
RT "Isolation, purification and cloning of a porcine mannose-binding
RT lectin.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF164576; AAD45377.1; -.
DR HSSP; P11226; 1HUP.
DR GO; GO:0005529; F:sugar binding; IEA.
DR GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00059; lectin_c; 1.
DR ProDom; PD000007; Clg_helix; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW Collagen; Lectin.
SQ SEQUENCE 240 AA; 25523 MW; 52BD865A21D3D563 CRC64;

Alignment Scores:

Pred. No.:	6.66e-68	Length:	240
Score:	808.00	Matches:	161
Percent Similarity:	74.90%	Conservative:	24
Best Local Similarity:	65.18%	Mismatches:	54
Query Match:	59.19%	Indels:	8
DB:	6	Gaps:	3

US-10-054-536-2 (1-747) x Q9XSW3 (1-240)

```
Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              ||||||||||||||||||||  ||||||||||||  ::::::  ||||||::::::
Db          1 MetSerLeuPheProSerLeuHisLeuLeuLeuLeuIleValMetThrAlaSerHisThr 20

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||  |||||||||  |||  ||||||  ||||||:::|||  |||
Db         21 GluThrGluAsnCysGluAspIleGlnAsnThrCys---LeuValIleSerCysAspSer 39

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||||||||||  ||||||||||||||  ||||||  ||||||||||||||
Db         40 ProGlyIleAsnGlyLeuProGlyLysAspGlyLeuAspGlyAlaLysGlyGluLysGly 59

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||  ||||||||||||  ||||||  :::||||||  |||
Db         60 GluProGlyGlnGlyLeuIleGlyLeuGlnGlyLeuProGlyMetValGlyProGlnGly 79

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              :::||||||  |||  ||||||  ||||||||||||||||||  |||
Db         80 SerProGlyIleProGlyLeuProGlyLeuLysGlyGlnLysGlyAspSerGlyIleAsp 99

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              |||  :::||||||||  |||::::::|||:::
Db        100 Pro------GlyAsnSerLeuAla--------------AsnLeuArgSerGluLeuAsp 112

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ||||||||||||||  |||:::  ||||||||||||||  |||  :::|||
Db        113 AsnIleLysLysTrpLeuIlePheAlaGlnGlyLysGlnValGlyLysLysLeuTyrLeu 132

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              |||||||||:::  |||:::|||  ||||||||||||||  :::||||||||||
Db        133 ThrAsnGlyLysLysMetSerPheAsnGlyValLysAlaLeuCysAlaGlnPheGlnAla 152

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              ||||||||||||||  |||:::  ||||||  ||||||||  |||
Db        153 SerValAlaThrProThrAsnSerArgGluAsnGlnAlaIleGlnGluLeuAlaGlyThr 172

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        173 GluAlaPheLeuGlyIleThrAspGluTyrThrGluGlyGlnPheValAspLeuThrGly 192

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
              |||:::  |||  |||||||||:::||||||||||||||||||  |||  |||
Db        193 LysArgValArgTyrGlnAsnTrpAsnAspGlyGluProAsnAsnAlaAspSerAlaGlu 212

Qy        661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
              ||||||  :::||||||:::|||:::|||||||||:::  ||||||:::
Db        213 HisCysValGluIleLeuLysAspGlyLysTrpAsnAspIlePheCysSerSerGlnLeu 232

Qy        721 CTGGCCGTCTGTGAGTTCCCT 741
              |||||||||||||||||||
Db        233 SerAlaValCysGluPhePro 239
```

RESULT 4
Q9Z294
ID Q9Z294 PRELIMINARY; PRT; 224 AA.
AC Q9Z294;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mannose-binding protein (Fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=86196130; PubMed=3009480;
RA Drickamer K., Dordal M.S., Reynolds L.;
RT "Mannose-binding proteins isolated from rat liver contain
RT carbohydrate-recognition domains linked to collagenous tails. Complete
RT primary structures and homology with pulmonary surfactant
RT apoprotein.";
RL J. Biol. Chem. 261:6878-6887(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Drickamer K., Dordal M.S., Reynolds L.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF080507; AAC31936.1; -.
DR HSSP; P19999; 1RTM.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF01391; Collagen; 1.

```
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   224 AA;  23800 MW;  44D790F8E12815A9 CRC64;

Alignment Scores:
Pred. No.:                3.87e-56      Length:          224
Score:                    684.00        Matches:         136
Percent Similarity:       69.83%        Conservative:    26
Best Local Similarity:    58.62%        Mismatches:      58
Query Match:              50.11%        Indels:          12
DB:                       11            Gaps:            3

US-10-054-536-2 (1-747) x Q9Z294 (1-224)

Qy         52 TCTTACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCC 111
              |||    |||    :::    |||||||||:::      |||||||||   :::||||||||| 
Db          2 SerSerSerGlySerGlnThrCysGluGluThrLeuLysThrCys---SerValIleAla 20

Qy        112 TGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGA 171
              |||                          |||:::|||||||||||||||   ||||||
Db         21 Cys--------------------------GlyArgAspGlyArgAspGlyProLysGly 31

Qy        172 GAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGG 231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         32 GluLysGlyGluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGly 51

Qy        232 CCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCT 291
              |||||||||:::   |||      ||||||  |||||||||||||||||||||||||
Db         52 ProProGlySerValGlyAlaProGlySerGlnGlyProLysGlyGlnLysGlyAspArg 71

Qy        292 GGAAAAAGTCCGGATGGTGATAGTAGCCTGGCT------GCCTCAGAAAGAAAAGCTCTG 345
              |||   |||          :::      ||||||        :::|||        |||
Db         72 GlyAspSerArgAlaIleGluValLysLeuAlaAsnMetGluAlaGluIleAsnThrLeu 91

Qy        346 CAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGG 405
              :::::::::::            |||       ||||||:::||||||:::   |||
Db         92 LysSerLysLeuGluLeuThrAsnLysLeuHisAlaPheSerMetGlyLysLysSerGly 111

Qy        406 AACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGT 465
              |||||||||:::||||||   |||   |||   |||||||||||||||||||
Db        112 LysLysPhePheValThrAsnHisGluArgMetProPheSerLysValLysAlaLeuCys 131

Qy        466 GTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAG 525
              :::   :::   :::||||||   ||||||||||   ||||||   |||||||||
Db        132 SerGluLeuArgGlyThrValAlaIleProArgAsnAlaGluGluAsnLysAlaIleGln 151

Qy        526 AATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTT 585
              :::   |||      |||||||||||||||||||||||||   ||||||||||||||||
Db        152 GluValAlaLysThrSerAlaPheLeuGlyIleThrAspGluValThrGluGlyGlnPhe 171

Qy        586 GTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAAT 645
              :::   :::||||||   |||||||||||:::||||||   :::   |||||||||:::
Db        172 MetTyrValThrGlyGlyArgLeuThrTyrSerAsnTrpLysLysAspGluProAsnAsp 191

Qy        646 GCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCC 705
              ||||||   ||||||||||   ::::::   ||||||   |||||||||:::
Db        192 HisGlySerGlyGluAspCysValThrIleValAspAsnGlyLeuTrpAsnAspIleSer 211

Qy        706 TGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCT 741
              |||      ||||||   ||||||||||||||||||
Db        212 CysGlnAlaSerHisThrAlaValCysGluPhePro 223

RESULT 5
Q7TMC0
ID   Q7TMC0       PRELIMINARY;      PRT;   371 AA.
AC   Q7TMC0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Ab2-011 (Ab2-001).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xu C.S., Li W.Q., Li Y.C., Han H.P., Wang G.P., Chai L.Q., Yuan J.Y.,
RA   Yang K.J., Yan H.M., Chang C.F., Zhao L.F., Ma H., Wang L., Wang S.F.,
RA   Shi J.B., Rahman S., Wang Q.N., Zhang J.B.;
RT   "Liver regeneration after PH.";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY325174; AAP92575.1; -.
DR   EMBL; AY325178; AAP92579.1; -.
SQ   SEQUENCE   371 AA;  39978 MW;  774F06369D895A03 CRC64;

Alignment Scores:
Pred. No.:                1.58e-51      Length:          371
Score:                    636.00        Matches:         126
Percent Similarity:       72.60%        Conservative:    33
Best Local Similarity:    57.53%        Mismatches:      56
Query Match:              46.59%        Indels:          4
DB:                       11            Gaps:            3

US-10-054-536-2 (1-747) x Q7TMC0 (1-371)

Qy          1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
              |||||||||||||    |||          |||||||||   ::::::   |||    |||::: 
Db          1 MetSerLeuPheThrSer------PheLeuLeuLeuCysValLeuThrAlaValTyrAla 18

Qy         61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
              ||||||:::|||   |||   ||||||   :::||||||   ||||||||||||||||||
Db         19 GluThrLeuThr---GluGlyAlaGlnSerSerCysPro---ValIleAlaCysSerSer 36

Qy        121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
              ||||||:::||||||||||||||||||||||||   ||||||   ||||||||||||||||
Db         37 ProGlyLeuAsnGlyPheProGlyLysAspGlyHisAspGlyAlaLysGlyGluLysGly 56

Qy        181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||   |||
Db         57 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysValGlyProAlaGly 76

Qy        241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
              ||||||   ||||||   |||   |||   |||||||||   |||:::|||
Db         77 ProProGlyAsnProGlySerLysGlyAlaThrGlyProLysGlyAspArgGlyGluSer 96

Qy        301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
              :::   |||::::::            |||      ||||||::::::|||:::
Db         97 ValGluPheAspThrThrAsnIleAspLeuGluIleAlaAlaLeuArgSerGluLeuArg 116

Qy        361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
              ::::::|||||:::      |||:::   :::   ||||||   |||:::|||:::
Db        117 AlaMetArgLysTrpValLeuLeuSerMetSerGluAsnValGlyLysLysTyrPheMet 136

Qy        421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
              ::::::          |||          :::   ||||||||||||   :::   |||
Db        137 SerSerValArgArgMetProLeuAsnArgAlaLysAlaLeuCysSerGluLeuGlnGly 156

Qy        481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
              :::|||||||||||||||||||||   ||||||   ||||||||||:::   |||:::
Db        157 ThrValAlaThrProArgAsnAlaGluGluAsnArgAlaIleGlnAsnValAlaLysAsp 176

Qy        541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
              ||||||||||||||||||||:::::|||||      |||   ||||||||||||
Db        177 ValAlaPheLeuGlyIleThrAspGlnArgThrGluAsnValPheGluAspLeuThrGly 196

Qy        601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
              ||||||:::   ||||||||||||||||||||      :::      |||   :::
Db        197 AsnArgValArgTyrThrAsnTrpAsnGluGlyArgLeuSerLeuArgGlyProGlu 215

RESULT 6
Q98TA4
ID   Q98TA4       PRELIMINARY;      PRT;   254 AA.
AC   Q98TA4;
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mannose-binding lectin protein precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20456722; PubMed=11003389;
RA   Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjodt K.;
RT   "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT   is expressed at high level in spleen, and the deduced primary
RT   structure predicts affinity for galactose.";
RL   Immunogenetics 51:955-964(2000).
DR   EMBL; AF231714; AAK30298.1; -.
DR   HSSP; P19999; 1YTT.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000005; HTHAraC.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
KW   Collagen; Lectin; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    254       MANNOSE-BINDING LECTIN PROTEIN.
FT   VARIANT     234    234       D -> V.
SQ   SEQUENCE   254 AA;  27376 MW;  C924428643441AED CRC64;

Alignment Scores:
Pred. No.:                1.51e-43      Length:          254
Score:                    551.50        Matches:         116
Percent Similarity:       64.40%        Conservative:    45
Best Local Similarity:    46.40%        Mismatches:      78
Query Match:              40.40%        Indels:          11
DB:                       13            Gaps:            6

US-10-054-536-2 (1-747) x Q98TA4 (1-254)

Qy          10 TTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTG 69
               |||   :::|||   |||   |||      |||:::|||   |||          |||
Db           7 PheSerAlaLeuLeuLeuCysLeuSerLeuMetMetAlaThrSerLeuLeuThrThrAsp 26

Qy          70 ACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATC 129
                  |||:::        :::||||||   :::|||   ||||||:::|||   :::
Db          27 LysProGluGluLysMetTyrSerCysPro---IleIleGlnCysSerAlaProAlaVal 45

Qy         130 AACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGC 189
               ||||||   ||||||:::|||||||||||||||   |||||||||||||:::||||||
Db          46 AsnGlyLeuProGlyArgAspGlyArgAspGlyProLysGlyGluLysGlyAspProGly 65

Qy         190 CAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGG 249
               :::||||||||||||||||||||   |||||||||   ||||||   |||      |||
Db          66 GluGlyLeuArgGlyLeuGlnGlyLeuProGlyLysAlaGlyProGlnGlyLeuLysGly 85

Qy         250 CCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGT 309
                  |||      |||   |||||||||||||||:::   |||
Db          86 GluValGlyProGlnGlyGluLysGlyGlnLysGlyGluArgGlyIleValValThrAsp 105

Qy         310 GAT---------------AGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
               |||               :::   |||   |||      |||      |||:::   :::

Db         106 AspLeuHisArgGlnIleThrAspLeuGluAlaLysIleArg---ValLeuGluAspAsp 124

Qy         355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
               ::::::|||   ||||||   |||:::                        ||||||   |||
Db         125 LeuSerArgTyrLysLysAlaLeuSerLeuLysAspValValAsnValGlyLysLysMet 144

Qy         415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
               |||::::::   |||:::         |||||||||   |||:::|||||||   |||
Db         145 PheValSerThrGlyLysLysTyrAsnPheGluLysGlyLysSerLeuCysAlaLysAla 164

Qy         475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
                  :::   :::|||:::|||||||||   |||||||||   |||:::::::::|||||
Db         165 GlySerValLeuAlaSerProArgAsnGluAlaGluAsnThrAlaLeuLysAspLeuIle 184

Qy         535 ------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTG 588
                        :::|||::::::||||||:::|||   :::|||||||||:::|||:::
Db         185 AspProSerSerGlnAlaTyrIleGlyIleSerAspAlaGlnThrGluGlyArgPheMet 204

Qy         589 GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT 648
                  |||:::|||      |||||||||:::||||||      ||||||||||||||
Db         205 TyrLeuSerGlyGlyProLeuThrTyrSerAsnTrpLysProGlyGluProAsnAsn--- 223

Qy         649 GGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGC 708
                     :::||||||||||   ::::::      :::|||:::||||||||:::   |||
Db         224 HisLysAsnGluAspCysAlaValIleGluAspSerGlyLysTrpAsnAspLeuAspCys 243

Qy         709 TCCACCTCCCATCTG---GCCGTCTGTGAG 735
               |||   |||::::::      :::||||||
Db         244 SerAsnSerAsnIlePheIleIleCysGlu 253

RESULT 7
O57451
ID   O57451      PRELIMINARY;      PRT;   238 AA.
AC   O57451;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mannan-binding lectin (Fragment).
GN   CMBL.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White Leghorn; TISSUE=Liver;
RA   Laursen S.B.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF022226; AAB94071.1; -.
DR   HSSP; P19999; 1YTT.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Collagen; Lectin.
FT   NON_TER     238    238
SQ   SEQUENCE   238 AA;  25645 MW;  E5C9B5197AAE64E3 CRC64;

Alignment Scores:
Pred. No.:                1.65e-42      Length:          238
Score:                    540.50        Matches:         111
Percent Similarity:       65.00%        Conservative:    45
Best Local Similarity:    46.25%        Mismatches:      73
Query Match:              39.60%        Indels:          11
DB:                       13            Gaps:            6
```

```
US-10-054-536-2 (1-747) x O57451 (1-238)

Qy          40 ATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCT 99
               |||:::|||   |||         |||         |||:::         :::|||||||
Db           1 MetMetAlaThrSerLeuLeuThrThrAspLysProGluGluLysMetTyrSerCysPro 20

Qy         100 GCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGAT 159
                  :::|||   ||||||:::|||   :::||||||   ||||||:::|||||||||||||
Db          21 ---IleIleGlnCysSerAlaProAlaValAsnGlyLeuProGlyArgAspGlyArgAsp 39

Qy         160 GGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCT 219
               |||   |||||||||||||||:::||||||:::|||||||||||||||||||||   |||
Db          40 GlyProLysGlyGluLysGlyAspProGlyGluGlyLeuArgGlyLeuGlnGlyLeuPro 59

Qy         220 GGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAA 279
               ||||||   ||||||   |||       |||       |||       |||   |||||||||
Db          60 GlyLysAlaGlyProGlnGlyLeuLysGlyGluValGlyProGlnGlyGluLysGlyGln 79

Qy         280 AAAGGAGACCCTGGAAAAAGTCCGGATGGTGAT--------------AGTAGCCTGGCT 324
               ||||||:::   |||                 |||             :::   |||
Db          80 LysGlyGluArgGlyIleValValThrAspAspLeuHisArgGlnIleThrAspLeuGlu 99

Qy         325 GCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTC 384
               |||       |||       |||:::   :::::::::|||   ||||||   |||:::
Db         100 AlaLysIleArg---ValLeuGluAspAspLeuSerArgTyrLysLysAlaLeuSerLeu 118

Qy         385 TCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTT 444
                        :::|||   |||   |||::::::   |||:::         |||
Db         119 LysAspValValAsnIleGlyLysLysMetPheValSerThrGlyLysLysTyrAsnPhe 138

Qy         445 GAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCT 504
               ||||||   |||:::||||||   |||       :::   :::|||:::|||||||||
Db         139 GluLysGlyLysSerLeuCysAlaLysAlaGlySerValLeuAlaSerProArgAsnGlu 158

Qy         505 GCAGAGAATGGAGCCATTCAGAATCTCATC------AAGGAGGAAGCCTTCCTGGGCATC 558
               |||||||||   |||:::::::::||||||           :::|||::::::||||||
Db         159 AlaGluAsnThrAlaLeuLysAspLeuIleAspProSerSerGlnAlaTyrIleGlyIle 178

Qy         559 ACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACA 618
               :::|||   :::|||||||||:::|||:::   |||:::|||       |||||||||:::
Db         179 SerAspAlaGlnThrGluGlyArgPheMetTyrLeuSerGlyGlyProLeuThrTyrSer 198

Qy         619 AACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTG 678
               ||||||      ||||||||||||||||           :::|||||||||   ::::::
Db         199 AsnTrpLysProGlyGluProAsnAsn---HisLysAsnGluAspCysAlaValIleGlu 217

Qy         679 AAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTG---GCCGTCTGTGAG 735
                  :::|||:::|||||||||:::   ||||||   |||:::::::      :::||||||
Db         218 AspSerGlyLysTrpAsnAspLeuAspCysSerAsnSerAsnIlePheIleIleCysGlu 237

RESULT 8
Q86YK9
ID   Q86YK9      PRELIMINARY;      PRT;   375 AA.
AC   Q86YK9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Surfactant, pulmonary-associated protein D.
GN   SFTPD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA   Ahearn M.O., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY216721; AAO22991.1; -.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   375 AA;  37698 MW;  79380764F2B86E67 CRC64;

Alignment Scores:
Pred. No.:                3e-34          Length:          375
Score:                    454.00         Matches:         99
Percent Similarity:       55.07%         Conservative:    26
Best Local Similarity:    43.61%         Mismatches:      82
Query Match:              33.26%         Indels:          20
DB:                       4              Gaps:            5

US-10-054-536-2 (1-747) x Q86YK9 (1-375)

Qy         118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
                  :::|||||||:::   |||       |||       |||       |||   |||||||||:::
Db         149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy         178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
               |||   ||||||:::           |||       |||       |||       |||
Db         169 GlyValProGlyGluArgGlyValProGlyAsnAlaGlyAlaAlaGlySerAlaGlyAla 188

Qy         226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
               :::||||||   |||:::||||||      |||   ||||||   ||||||   ||||||
Db         189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy         286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
                  ||||||   |||      |||:::|||   |||       ||||||   |||
Db         209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy         337 AAAGCT------------------------------CTGCAAACAGAAATGGCACGTATC 366
                  :::                             ||||||      ::::::
Db         229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy         367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
               ||||||          |||       |||:::   ||||||   |||   |||   |||
Db         249 LysLysValGluLeuProPheProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268

Qy         427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
               |||   :::       |||   :::   :::   ||||||   :::           :::
Db         269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy         487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
               |||:::||||||:::|||||||||||   |||:::|||   |||:::
Db         289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy         538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
               |||   |||||||||   :::||||||   ||||||||||:::|||           |||
Db         309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy         598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
               |||       |||   |||:::||||||      ||||||||||||:::   |||
Db         329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyGluProAsnAspAspGlyGlySer 348

Qy         658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
               ||||||||||||   :::      ||||||:::|||||||||      |||
Db         349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy         718 CATCTGGCCGTCTGTGAGTTC 738
                  |||   ||||||||||||
Db         369 ArgLeuValValCysGluPhe 375

RESULT 9
Q8TCD8
ID   Q8TCD8      PRELIMINARY;      PRT;   375 AA.
AC   Q8TCD8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC022318; AAH22318.1; -.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 4.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Hypothetical protein; Collagen.
SQ   SEQUENCE   375 AA;  37655 MW;  CCB7375D3C86421A CRC64;

Alignment Scores:
Pred. No.:                7.19e-34      Length:          375
Score:                    450.00        Matches:         98
Percent Similarity:       55.07%        Conservative:    27
Best Local Similarity:    43.17%        Mismatches:      82
Query Match:              32.97%        Indels:          20
DB:                       4             Gaps:            5

US-10-054-536-2 (1-747) x Q8TCD8 (1-375)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||:::   |||      |||      |||      |||   |||||||||:::
Db        149 AlaProGlyMetGlnGlySerAlaGlyAlaArgGlyLeuAlaGlyProLysGlyGluArg 168

Qy        178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
              |||   ||||||:::           |||      |||      |||      |||
Db        169 GlyValProGlyGluArgGlyValProGlyAsnThrGlyAlaAlaGlySerAlaGlyAla 188

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||:::||||||     |||  ||||||  ||||||  ||||||
Db        189 MetGlyProGlnGlySerProGlyAlaArgGlyProProGlyLeuLysGlyAspLysGly 208

Qy        286 GACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTG------GCTGCCTCAGAAAGA 336
                  ||||||   |||          |||:::|||   |||          ||||||   |||
Db        209 IleProGlyAspLysGlyAlaLysGlyGluSerGlyLeuProAspValAlaSerLeuArg 228

Qy        337 AAAGCT----------------------------CTGCAAACAGAAATGGCACGTATC 366
              :::                              ||||||        ::::::
Db        229 GlnGlnValGluAlaLeuGlnGlyGlnValGlnHisLeuGlnAlaAlaPheSerGlnTyr 248

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||        |||      |||:::   ||||||  |||   |||   |||
Db        249 LysLysValGluLeuProProAsnGlyGlnSerValGlyGluLysIlePheLysThrAla 268

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::     |||  :::   :::   ||||||  :::          :::
Db        269 GlyPheValLysProPheThrGluAlaGlnLeuLeuCysThrGlnAlaGlyGlyGlnLeu 288

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAG 537
              |||:::||||||:::||||||||||||   |||:::|||   |||:::
Db        289 AlaSerProArgSerAlaAlaGluAsnAlaAlaLeuGlnGlnLeuValValAlaLysAsn 308

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              |||   |||||||||   :::||||||   ||||||||||||:::|||         |||
Db        309 GluAlaAlaPheLeuSerMetThrAspSerLysThrGluGlyLysPheThrTyrProThr 328

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGAT 657
              |||      |||   |||:::||||||      |||:::||||||:::   |||
Db        329 GlyGluSerLeuValTyrSerAsnTrpAlaProGlyLysProAsnAspAspGlyGlySer 348

Qy        658 GAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC 717
              ||||||||||   :::      ||||||:::|||||||||      |||
Db        349 GluAspCysValGluIlePheThrAsnGlyLysTrpAsnAspArgAlaCysGlyGluLys 368

Qy        718 CATCTGGCCGTCTGTGAGTTC 738
                  |||   ||||||||||||
Db        369 ArgLeuValValCysGluPhe 375

RESULT 10
Q9I9Q7
ID   Q9I9Q7      PRELIMINARY;      PRT;   246 AA.
AC   Q9I9Q7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mannose binding-like lectin precursor (Fragment).
GN   MBL.
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Carassius.
OX   NCBI_TaxID=7957;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20456722; PubMed=11003389;
RA   Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjodt K.;
RT   "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT   is expressed at high level in spleen, and the deduced primary
RT   structure predicts affinity for galactose.";
RL   Immunogenetics 51:955-964(2000).
DR   EMBL; AF227739; AAF63470.1; -.
DR   HSSP; P35247; 1B08.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Lectin; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     13       POTENTIAL.
FT   VARIANT     145    145       S -> F.
SQ   SEQUENCE   246 AA;  25709 MW;  AB692282D289D0D5 CRC64;

Alignment Scores:
Pred. No.:                7.94e-32      Length:          246
Score:                    428.00        Matches:         89
Percent Similarity:       55.93%        Conservative:    43
Best Local Similarity:    37.71%        Mismatches:      86
Query Match:              31.36%        Indels:          18
DB:                       13            Gaps:            5

US-10-054-536-2 (1-747) x Q9I9Q7 (1-246)

Qy         76 GAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGT---AGCTCTCCAGGCATCAAC 132
              |||                |||||||||                  :::||||||   |||
Db         15 GluProGlnAsnLeuAsnCysProAlaTyrGlyGlyValProGlyThrProGlyHisAsn 34

Qy        133 GGCTTCCCAGGCAAAGATGGGCGTGAT-----------------GGCACCAAGGGAGAA 174
              |||   ||||||:::||||||||||||                 |||   |||||||||
Db         35 GlyLeuProGlyArgAspGlyArgAspGlyLysAspGlyAlaIleGlyProLysGlyGlu 54

Qy        175 AAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCT 234
              |||||||||   |||                :::|||||||||||||||||   ||||||
Db         55 LysGlyGluSerGlyVal---------SerValGlnGlyProProGlyLysAlaGlyPro 71

Qy        235 CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA 294
              ||||||      |||      |||      ||||||   |||   :::|||   ||||||
```

```
Db         72 ProGlyThrAlaGlyGluLysGlyGluArgGlyProSerGlyProGlnGlySerProGly 91

Qy        295 AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
                           :::|||            ||||||  :::   |||:::  :::
Db         92 SerGluSerValLeuGluSer------LeuLysSerGluIleGlnGlnLeuLysAlaLys 109

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
              :::|||       :::|||      :::          ::::::||||||   |||:::
Db        110 IleAlaThrPheGluLysValSerSerValCysHisPheArgLysValGlyGlnLysTyr 129

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              ::::::|||:::|||   :::       |||::::::            |||::::::|||
Db        130 TyrIleThrAspGlyValValGlyAsnPheAspGlnGlyLeuLysSerCysMetGluPhe 149

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
                 ::::::   :::||||||   :::|||||||||   |||:::      |||:::
Db        150 GlyGlyThrMetValSerProArgThrSerAlaGluAsnGlnAlaLeuLeuLysLeuVal 169

Qy        535 ------------------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAA 576
                                ::::::   ::::::|||:::||||||   |||||||||
Db        170 ValSerSerGlyLeuGlySerLysLysProTyrIleGlyValThrAspArgLysThrGlu 189

Qy        577 GGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAA 636
              |||||||||||||||      |||   :::||||||:::|||||||||      |||:::
Db        190 GlyGlnPheValAspThrGluGlyLysGlnLeuThrPheThrAsnTrpGlyProGlyGln 209

Qy        637 CCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAAT 696
              |||::::::            :::||||||   ::::::         |||   |||:::
Db        210 ProAspAspTyrLysGlyLeuGlnAspCysGlyValIleGluAspThrGlyLeuTrpAsp 229

Qy        697 GACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATC 744
              |||      |||                 :::||||||     |||
Db        230 AspGlyGlyCysGlyAspIleArgProIleMetCysGluIleAspIle 245
```

RESULT 11
Q863A1

```
ID   Q863A1      PRELIMINARY;      PRT;   369 AA.
AC   Q863A1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Surfactant protein D precursor.
GN   SP-D.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gjerstorff M., Hansen S., Madsen J., Bendixen C., Holmskov U.;
RT   "Bovine surfactant protein D: Genomic characterization, chromosomal
RT   localization and tissue distribution.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ548848; CAD69922.1; -.
DR   EMBL; AJ548849; CAD69922.1; JOINED.
DR   EMBL; AJ548850; CAD69922.1; JOINED.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Signal.
FT   SIGNAL        1     20       Potential.
FT   CHAIN        21     67       surfactant protein D.
FT   CHAIN       245    369       surfactant protein D.
SQ   SEQUENCE   369 AA;  37405 MW;  4A74B7593508AE5D CRC64;

Alignment Scores:
Pred. No.:              2.92e-31      Length:        369
Score:                  422.50        Matches:          88
Percent Similarity:     52.68%        Conservative:     30
Best Local Similarity:  39.29%        Mismatches:       89
Query Match:            30.95%        Indels:           17
DB:                     6             Gaps:              4
```

US-10-054-536-2 (1-747) x Q863A1 (1-369)

```
Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::|||||||||   |||   ||||||         |||         |||   :::|||
Db        146 AlaProGlyIleGlnGlySerProGlyProAlaGlyLeuLysGlyGluArgGlyAlaPro 165

Qy        178 GGGGAACCAGGC---CAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCT 234
              ||||||||||||      |||      |||      ||||||   |||   :::||||||
Db        166 GlyGluProGlyAlaProGlyArgAlaGlyAlaProGlyProAlaGlyAlaIleGlyPro 185

Qy        235 CCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA 294
                 |||      |||      |||   ||||||   ||||||   :::|||   ||||||
Db        186 GlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGlyThrProGly 205

Qy        295 ---AAAAGTCCGGATGGTGATAGTAGCCTGGCT--------------------------- 324
                 :::         |||:::|||   ||||||
Db        206 GluArgGlyAlaLysGlyGluSerGlyLeuAlaGluValAsnAlaLeuArgGlnArgVal 225

Qy        325 ---------GCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGG 375
                          :::   :::   ||||||         ::::::   ||||||
Db        226 GlyIleLeuGluGlyGlnLeuGlnArgLeuGlnAsnAlaPheSerGlnTyrLysLysAla 245

Qy        376 CTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATA 435
              :::   |||      |||:::   ||||||   |||   |||   |||   |||
Db        246 MetLeuPheProAsnGlyArgSerValGlyGluLysIlePheLysThrGluGlySerGlu 265

Qy        436 ATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCC 495
                 ||||||:::      :::   :::|||   :::               :::   :::|||
Db        266 LysThrPheGlnAspAlaGlnGlnIleCysThrGlnAlaGlyGlyGlnLeuProSerPro 285

Qy        496 AGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---------AAGGAGGAAGCC 546
              |||:::||||||||||||   |||:::         |||            :::   |||
Db        286 ArgSerAlaAlaGluAsnGluAlaLeuThrGlnLeuAlaThrAlaGlnAsnLysAlaAla 305

Qy        547 TTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGA 606
              ||||||   ::::::|||   :::   ||||||   |||:::         ||||||
Db        306 PheLeuSerMetSerAspThrArgLysGluGlyThrPheIleTyrProThrGlyGluPro 325

Qy        607 CTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGT 666
              |||   |||:::||||||         ||||||||||||   |||      |||:::|||
Db        326 LeuValTyrSerAsnTrpAlaProGlnGluProAsnAsnAspGlyGlySerGluAsnCys 345

Qy        667 GTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCC 726
              |||   :::         ||||||:::|||||||||         |||         |||
Db        346 ValGluIlePheProAsnGlyLysTrpAsnAspLysValCysGlyGluGlnArgLeuVal 365

Qy        727 GTCTGTGAGTTC 738
              :::|||||||||
Db        366 IleCysGluPhe 369
```

RESULT 12
Q9TT06

```
ID   Q9TT06      PRELIMINARY;      PRT;   248 AA.
AC   Q9TT06;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Pulmonary surfactant protein A (Pulmonary surfactant-associated
DE   protein A).
GN   SPAS OR SP-A.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=20215262; PubMed=10749753;
RA   Braems G.A., Yao L.-J., Inchley K., Brickenden A., Han V.K.M.,
RA   Grolla A., Challis J.R.G., Possmayer F.;
RT   "Ovine surfactant protein cDNAs: use in studies on fetal lung growth
RT   and maturation after prolonged hypoxemia.";
RL   Am. J. Physiol. 278:L754-L764(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=20215263; PubMed=10749754;
RA   Pietschmann S.M., Pison U.;
RT   "cDNA cloning of ovine pulmonary SP-A, SP-B, and SP-C: isolation of
RT   two different sequences for SP-B.";
RL   Am. J. Physiol. 278:L765-L778(2000).
DR   EMBL; AF211856; AAF18995.1; -.
DR   EMBL; AF076633; AAF31148.1; -.
DR   HSSP; P22897; 1EGG.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   248 AA;  26394 MW;  D65E7293BBFF1FD9 CRC64;
```

Alignment Scores:			
Pred. No.:	5.1e-31	Length:	248
Score:	419.50	Matches:	102
Percent Similarity:	52.45%	Conservative:	37
Best Local Similarity:	38.49%	Mismatches:	87
Query Match:	30.73%	Indels:	39
DB:	6	Gaps:	12

US-10-054-536-2 (1-747) x Q9TT06 (1-248)

```
Qy          7 CTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACT 66
              |||      ||||||   |||:::||||||   |||||||||:::
Db          2 LeuLeuCysSerLeuThrLeuMetLeuLeuTrpMetValAlaSerGly------------ 17

Qy         67 GTGACCTGTGAGGATGCCCAAAAGACCTGC---------CCTGCAGTGATTGCCTGTAGC 117
              :::   |||   |||   ::::::   |||         |||   :::
Db         18 LeuGluCys---AspThrLysGluValCysLeuGlySerProGlyIle------ProGly 34

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::||||||   :::|||   ||||||:::|||||||||||||||   ||||||:::
Db         35 ThrProGlySerHisGlyLeuProGlyArgAspGlyArgAspGlyIleLysGlyAspPro 54

Qy        178 GGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAGTTGGGGCCTCCA 237
              |||   ||||||                        ||||||||||||   |||   |||
Db         55 GlyProProGlyPro--------------MetGlyProProGly------GlyMetPro 67

Qy        238 GGAAATCCAGGGCCTTCTGGGTCA-----------------CCAGGACCAAAGGGCCAA 279
              |||   ||||||   |||                      ||||||   :::|||:::
Db         68 GlyLeuProGlyArgAspGlyMetThrGlyAlaProGlyLeuProGlyGluArgGlyGlu 87

Qy        280 AAAGGAGACCCTGGA---AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGA 336
              ||||||:::||||||   :::   |||   |||      :::                |||
Db         88 LysGlyGluProGlyGluArgGlyProProGlyPheProAlaTyrLeuAspGluGlu--- 106

Qy        337 AAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGC--- 393
              ||||||   :::   |||:::   :::   |||   |||
Db        107 ------LeuGlnGlyThrLeuHisGluIleArgHisGlnValLeuGlnSerGlnGlyVal 124

Qy        394 --------------------AAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAA 432
                                  :::||||||   |||   |||   |||||||||:::
Db        125 LeuIleLeuGlnGlySerMetLeuGluValGlyGluLysValPheSerThrAsnGlyGln 144

Qy        433 ATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACC 492
              :::   |||:::   :::|||   ||||||   :::                :::|||
Db        145 SerLeuAsnPheAspAlaIleLysGluLeuCysAlaArgAlaGlyGlyHisIleAlaAla 164

Qy        493 CCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAGGAA--------- 543
              ||||||:::      ||||||   ||||||   :::::::::|||:::
Db        165 ProArgSerProGluGluAsnGluAlaIleThrSerIleValLysLysHisAsnThrTyr 184

Qy        544 GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAAT 603
              |||:::||||||:::   :::      |||   |||   |||      |||   |||
Db        185 AlaTyrLeuGlyLeuAlaGluGlyProThrAlaGlyAspPheTyrTyrLeuAspGlyAla 204

Qy        604 AGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGAT 663
                  :::   ||||||||||||      |||||||||          |||      |||
Db        205 ProValAsnTyrThrAsnTrpTyrProGlyGluProArgGlyArgGlyLys---GluLys 223

Qy        664 TGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTG 723
              ||||||   :::      :::||||||||||||||||      |||            |||
Db        224 CysValGluIleTyrThrAspGlyGlnTrpAsnAspLysAsnCysLeuGlnTyrArgLeu 243

Qy        724 GCCGTCTGTGAGTTC 738
              |||:::|||||||||
Db        244 AlaIleCysGluPhe 248
```

```
RESULT 13
Q9I9Q9
ID   Q9I9Q9      PRELIMINARY;      PRT;   256 AA.
AC   Q9I9Q9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mannose binding-like lectin precursor.
GN   MBL.
OS   Cyprinus carpio (Common carp).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Cyprinus.
OX   NCBI_TaxID=7962;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20456722; PubMed=11003389;
RA   Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjodt K.;
RT   "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT   is expressed at high level in spleen, and the deduced primary
RT   structure predicts affinity for galactose.";
RL   Immunogenetics 51:955-964(2000).
DR   EMBL; AF227737; AAF63468.1; -.
DR   HSSP; P35247; 1B08.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Lectin; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    256       MANNOSE BINDING-LIKE LECTIN.
FT   VARIANT     235    235       S -> T.
SQ   SEQUENCE   256 AA;  26934 MW;  D019291D1167730D CRC64;
```

Alignment Scores:			
Pred. No.:	6.4e-31	Length:	256
Score:	418.50	Matches:	95
Percent Similarity:	54.05%	Conservative:	45
Best Local Similarity:	36.68%	Mismatches:	98
Query Match:	30.66%	Indels:	21
DB:	13	Gaps:	7

US-10-054-536-2 (1-747) x Q9I9Q9 (1-256)

```
Qy          1 ATGTCCCTGTTTCCA---TCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTAC 57
              |||:::||||||        |||      |||||||||        |||
Db          1 MetAlaLeuPheLysLeuPheLeuGlyThrLeuLeuLeuLeuGlnPheAlaLeuGlnLeu 20
```

```
Qy        58 TCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGT--- 114
Db        21 LeuAsp------GlyAlaGluProGlnAsnLeuAsnCysProAlaTyrGlyGlyValPro 38

Qy       115 AGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGAT-------------- 159
Db        39 GlyThrProGlyHisAsnGlyLeuProGlyArgAspGlyArgAspGlyLysAspGlyAla 58

Qy       160 ---GGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCC 216
Db        59 IleGlyProLysGlyGluLysGlyGluSerGlyVal---------SerValGlnGlyPro 75

Qy       217 CCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGC 276
Db        76 ProGlyLysAlaGlyProProGlyProAlaGlyGluLysGlyGluArgGlyProThrGly 95

Qy       277 CAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGA 336
Db        96 SerGlnGlySerProGlySerGluSerValLeuGluSer------LeuLysSerGluIle 113

Qy       337 AAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAA 396
Db       114 GlnGlnLeuLysAlaLysIleAlaThrPheGluLysValAlaSerValGlyHisPheArg 133

Qy       397 CAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAG 456
Db       134 GlnValGlyGlnLysTyrTyrIleThrAspGlyValValGlyThrPheAspGlnGlyLeu 153

Qy       457 GCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGA 516
Db       154 LysPheCysLysAspPheGlyGlyThrMetValPheProArgThrSerAlaGluAsnGln 173

Qy       517 GCCATTCAGAATCTCATC----------------AAGGAGGAAGCCTTCCTGGGCATC 558
Db       174 AlaLeuLeuLysLeuValValSerSerGlyLeuSerSerLysLysProTyrIleGlyVal 193

Qy       559 ACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACA 618
Db       194 ThrAspArgGluThrGluGlyArgPheValAsnThrGluGlyLysGlnLeuThrPheThr 213

Qy       619 AACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTG 678
Db       214 AsnTrpGlyProGlyGlnProAspAspTyrLysGlyLeuGlnAspCysGlyValIleGlu 233

Qy       679 AAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAG 735
Db       234 AspSerGlyLeuTrpAspAspGlySerCysGlyAspIleArgProIleMetCysGlu 252
```

```
RESULT 14
Q9N1X4
ID   Q9N1X4      PRELIMINARY;      PRT;   378 AA.
AC   Q9N1X4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lung surfactant protein D precursor.
GN   SFTPD.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20109098; PubMed=10640760;
RA   van Eijk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA   Lawson P.R.;
RT   "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT   localisation and tissue distribution.";
RL   J. Immunol. 164:1442-1450(2000).
DR   EMBL; AF132496; AAF22145.2; -.
DR   HSSP; P35247; 1B08.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 3.
DR   Pfam; PF00059; lectin_c; 1.
DR   ProDom; PD000007; Clg_helix; 2.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Collagen; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    378       LUNG SURFACTANT PROTEIN D.
SQ   SEQUENCE   378 AA;  37986 MW;  3504E8C1E56C341D CRC64;
```

```
Alignment Scores:
Pred. No.:                3.63e-30        Length:          378
Score:                    411.00          Matches:         110
Percent Similarity:       37.89%          Conservative:     34
Best Local Similarity:    28.95%          Mismatches:      100
Query Match:              30.11%          Indels:          136
DB:                       6               Gaps:             11

US-10-054-536-2 (1-747) x Q9N1X4 (1-378)

Qy         1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
Db         1 MetLeuLeuLeuProLeuSerValLeuIleLeuLeuThrGlnProProArgSerLeuGly 20

Qy        61 ---------GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCC 111
Db        21 AlaGluMetLysThrTyrSerGlnArgAlaValAlaAsnAlaCys---AlaLeuValMet 39

Qy       112 TGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGAT---------GGC 162
Db        40 Cys---SerProMetGluAsnGlyLeuProGlyArgAspGlyArgAspGlyArgGluGly 58

Qy       163 ACCAAGGGAGAAAAGGGGGAACCAGGC---------------------CAAGGGCTCAGA 201
Db        59 ProArgGlyGluLysGlyAspProGlyLeuProGlyAlaValGlyArgAlaGlyMetPro 78

Qy       202 GGCTTACAGGGCCCC---------------------------------------------- 216
Db        79 GlyLeuAlaGlyProValGlyProLysGlyAspAsnGlySerThrGlyGluProGlyAla 98

Qy       216 ---------------------------------------------------------- 216
Db        99 LysGlyAspIleGlyProCysGlyProProGlyProProGlyIleProGlyProAlaGly 118

Qy       217 ---------------------CCTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCT 255
Db       119 LysGluGlyProSerGlyGlnGlnGlyAsnIleGlyProProGlyThrProGlyProLys 138

Qy       256 GGGTCACCAGGACCAAAGGGCCAA---------------------------------- 279
Db       139 GlyGluThrGlyProLysGlyGluValGlyAlaLeuGlyMetGlnGlySerThrGlyAla 158

Qy       279 ---------------------------------------------------------- 279
Db       159 ArgGlyProAlaGlyLeuLysGlyGluArgGlyAlaProGlyGluArgGlyAlaProGly 178

Qy       280 ----------------------------------------------------------AAA 282
Db       179 SerAlaGlyAlaAlaGlyProAlaGlyAlaThrGlyProGlnGlyProSerGlyAlaArg 198

Qy       283 GGAGACCCTGGAAAAAGTCCGGAT----------------------------GGTGAT 312
Db       199 GlyProProGlyLeuLysGlyAspArgGlyProProGlyGluArgGlyAlaLysGlyGlu 218

Qy       313 AGTAGCCTGGCTGCC---------------------------------TCAGAAAGA 336
Db       219 SerGlyLeuProGlyIleThrAlaLeuArgGlnGlnValGluThrLeuGlnGlyGlnVal 238

Qy       337 AAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAA 396
```

```
                 :::  ||||||        ::::::  ||||||          |||     |||:::
Db       239 GlnArgLeuGlnLysAlaPheSerGlnTyrLysLysValGluLeuPheProAsnGlyArg 258

Qy       397 CAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAG 456
                 ||||||   |||   |||   |||   |||          ||||||:::      :::
Db       259 GlyValGlyGluLysIlePheLysThrGlyGlyPheGluLysThrPheGlnAspAlaGln 278

Qy       457 GCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGA 516
                 :::|||   :::               :::|||:::||||||:::      ||||||
Db       279 GlnValCysThrGlnAlaGlyGlyGlnMetAlaSerProArgSerGluThrGluAsnGlu 298

Qy       517 GCCATTCAGAATCTCATC---------AAGGAGGAAGCCTTCCTGGGCATCACTGATGAG 567
             |||:::      |||:::               :::   |||||||||   :::||||||
Db       299 AlaLeuSerGlnLeuValThrAlaGlnAsnLysAlaAlaPheLeuSerMetThrAspIle 318

Qy       568 AAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAAC 627
             ||||||||||||   |||            ||||||      |||   |||   ||||||
Db       319 LysThrGluGlyAsnPheThrTyrProThrGlyGluProLeuValTyrAlaAsnTrpAla 338

Qy       628 GAGGGTGAACCCAACAATGCTGGTTCTGAT---------GAAGATTGTGTATTGCTACTG 678
                ||||||||||||||   |||               |||:::||||||   :::
Db       339 ProGlyGluProAsnAsnAsnGlyGlySerSerGlyAlaGluAsnCysValGluIlePhe 358

Qy       679 AAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
                ||||||:::|||||||||         |||            |||   :::|||||||||
Db       359 ProAsnGlyLysTrpAsnAspLysAlaCysGlyGluLeuArgLeuValIleCysGluPhe 378
```

RESULT 15
Q95L88

```
ID   Q95L88      PRELIMINARY;      PRT;   248 AA.
AC   Q95L88;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Pulmonary surfactant-associated protein A.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Weber B.I.L., Hospes R., Gortner L.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF400580; AAL07690.1; -.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   248 AA;  26047 MW;  B71133E005C9A5C1 CRC64;

Alignment Scores:
Pred. No.:              7.84e-30      Length:         248
Score:                  407.00        Matches:        97
Percent Similarity:     51.74%        Conservative:   37
Best Local Similarity:  37.45%        Mismatches:     87
Query Match:            29.82%        Indels:         38
DB:                     6             Gaps:           12

US-10-054-536-2 (1-747) x Q95L88 (1-248)

Qy        28 CTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCCAA 87
             |||   |||:::::::::      :::   |||   |||      |||   |||   :::
Db         6 LeuThrLeuThrLeuIleLeuLeuAlaValSerGlyThr---LysCys---AspValLys 23

Qy        88 AAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT------------CCAGGCATCAACGGC 135
             :::   |||               |||||||||             ||||||   |||
Db        24 GluPheCys---------AlaAlaCysSerGlyValProGlyIleProGlySerProGly 40

Qy       136 TTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGGAACCAGGCCAAGGG 195
             ||||||:::||||||||||||||   ||||||:::   |||   ||||||
Db        41 LeuProGlyArgAspGlyArgAspGlyValLysGlyAspProGlyProProGlyPro--- 59

Qy       196 CTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGAAATCCA--------- 246
                         ||||||||||||      |||   ||||||:::|||
Db        60 ------------IleGlyProProGly------GlyMetProGlySerProGlyHisAsp 73

Qy       247 ---------GGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGA--- 294
                      ||||||   |||   ||||||   :::|||   ||||||:::||||||
Db        74 GlyLeuIleGlyProProGlyProProGlyGluArgGlyAspLysGlyGluProGlyGlu 93

Qy       295 AAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
             :::   |||   |||      :::               |||         |||||||||
Db        94 ArgGlyProProGlyProProAlaTyrProAspGluGlu---------LeuGlnThrThr 110

Qy       355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAA------------------ 396
             :::      |||:::      :::         :::|||
Db       111 LeuHisAspIleArgHisGlnIleLeuGlnLeuMetGlyAlaLeuSerLeuGlnGlySer 130

Qy       397 ------CAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAA 450
                   ||||||   |||   |||   |||||||||:::::::::   |||:::
Db       131 MetLeuAlaValGlyGluLysValPheSerThrAsnGlyGlnValValAspPheAspAla 150

Qy       451 GTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAG 510
             ::::::      |||   :::            :::|||   |||::::::      |||
Db       151 IleArgGluSerCysAlaArgAlaGlyGlyArgIleAlaValProLysSerLeuGluGlu 170

Qy       511 AATGGAGCCATTCAGAATCTCATCAAGGAGGAA---------GCCTTCCTGGGCATCACT 561
             |||   ||||||   :::|||:::   :::               |||:::||||||:::
Db       171 AsnAlaAlaIleAlaSerLeuValThrLysHisAsnThrTyrAlaTyrLeuGlyLeuGlu 190

Qy       562 GATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAAC 621
             :::      |||   |||   |||         |||   |||      :::   |||||||||
Db       191 GluGlyProThrAlaGlyAspPheTyrTyrLeuAspGlyAlaProValAsnTyrThrAsn 210

Qy       622 TGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAA 681
             |||      |||||||||            |||      |||   ||||||   :::
Db       211 TrpTyrProGlyGluProArgGlyArgGlyLys---GluLysCysValGluMetTyrThr 229

Qy       682 AATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
             :::|||||||||||||||         |||               ||||||:::|||||||||
Db       230 AspGlyGlnTrpAsnAspArgSerCysLeuGlnTyrArgLeuAlaIleCysGluPhe 248

Search completed: June 18, 2004, 21:02:45
Job time : 54.5 secs
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 18, 2004, 20:58:07 ; Search time 20.5 Seconds
(without alignments)
3762.400 Million cell updates/sec

Title: US-10-054-536-2
Perfect score: 1365
Sequence: 1 atgtccctgtttccatcact.....tctgtgagttccctatctga 747

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10054536/runat_18062004_173456_13622/app_query.fasta_1.903
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054536_@CGN_1_1_27_@runat_18062004_173456_13622 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1321	96.8	248	4	US-09-198-603C-2	Sequence 2, Appli
2	442.5	32.4	351	3	US-09-011-735-1	Sequence 1, Appli
3	442.5	32.4	351	4	US-09-029-156-1	Sequence 1, Appli
4	359	26.3	108	6	5514582-16	Patent No. 5514582
5	347.5	25.5	107	6	5514582-17	Patent No. 5514582
6	286.5	21.0	161	3	US-09-011-735-6	Sequence 6, Appli
7	212.5	15.6	111	6	5514582-20	Patent No. 5514582
8	204	14.9	197	4	US-09-602-877A-99	Sequence 99, Appl
9	201.5	14.8	287	1	US-08-365-103B-4	Sequence 4, Appli
10	201.5	14.8	300	1	US-08-365-103B-6	Sequence 6, Appli
11	201.5	14.8	327	1	US-08-365-103B-2	Sequence 2, Appli
12	196.5	14.4	208	4	US-09-535-521-20	Sequence 20, Appl
13	196.5	14.4	292	4	US-09-535-521-2	Sequence 2, Appli
14	196.5	14.4	292	4	US-09-535-521-5	Sequence 5, Appli
15	196	14.4	187	4	US-09-535-521-17	Sequence 17, Appl
16	193	14.1	128	4	US-09-535-521-8	Sequence 8, Appli
17	193	14.1	139	4	US-09-535-521-11	Sequence 11, Appl
18	193	14.1	141	4	US-09-535-521-14	Sequence 14, Appl
19	192	14.1	244	2	US-08-463-911-7	Sequence 7, Appli
20	192	14.1	244	3	US-09-140-804-3	Sequence 3, Appli
21	192	14.1	244	4	US-09-336-536-20	Sequence 20, Appl
22	192	14.1	244	4	US-09-530-423-1	Sequence 1, Appli
23	192	14.1	244	4	US-09-686-838B-3	Sequence 3, Appli
24	192	14.1	244	4	US-09-911-176B-48	Sequence 48, Appl
25	192	14.1	244	4	US-09-552-225A-3	Sequence 3, Appli
26	192	14.1	244	4	US-09-619-740-51	Sequence 51, Appl
27	192	14.1	244	4	US-09-776-976-6	Sequence 6, Appli
28	192	14.1	244	4	US-09-909-547-6	Sequence 6, Appli
29	192	14.1	244	4	US-09-569-852B-6	Sequence 6, Appli
30	192	14.1	244	4	US-09-552-204A-3	Sequence 3, Appli
31	189.5	13.9	123	4	US-09-535-521-25	Sequence 25, Appl
32	189.5	13.9	247	2	US-08-463-911-2	Sequence 2, Appli
33	189.5	13.9	247	4	US-09-776-976-4	Sequence 4, Appli
34	189.5	13.9	247	4	US-09-909-547-4	Sequence 4, Appli
35	189	13.8	125	3	US-08-722-126A-7	Sequence 7, Appli
36	189	13.8	125	5	PCT-US95-04258-7	Sequence 7, Appli
37	188	13.8	231	4	US-09-530-423-2	Sequence 2, Appli
38	187.5	13.7	280	4	US-09-247-155-178	Sequence 178, App
39	187	13.7	285	4	US-09-312-283C-382	Sequence 382, App
40	186.5	13.7	111	6	5514582-11	Patent No. 5514582
41	186.5	13.7	245	4	US-09-552-225A-4	Sequence 4, Appli
42	186.5	13.7	245	4	US-09-552-204A-4	Sequence 4, Appli
43	185.5	13.6	202	1	US-08-469-486-56	Sequence 56, Appl
44	185.5	13.6	202	2	US-08-469-658-56	Sequence 56, Appl
45	185	13.6	912	5	PCT-US95-03747-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-198-603C-2
; Sequence 2, Application US/09198603C
; Patent No. 6337193
; GENERAL INFORMATION:
; APPLICANT: TULLY, Raymond E.
; APPLICANT: CALTAGIRONE, G. Thomas
; APPLICANT: MOYER, Shawn S.
; APPLICANT: RONNING, Michael T.
; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC
; TITLE OF INVENTION: YEAST
; FILE REFERENCE: A7290
; CURRENT APPLICATION NUMBER: US/09/198,603C
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Human
US-09-198-603C-2

Alignment Scores:

Pred. No.:	2.52e-126	Length:	248
Score:	1321.00	Matches:	248
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.78%	Indels:	0
DB:	4	Gaps:	0

US-10-054-536-2 (1-747) x US-09-198-603C-2 (1-248)

Qy 1 ATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCA 60
Db 1 MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAlaAlaSerTyrSer 20

```
Qy        61 GAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 GluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer  40

Qy       121 CCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGly  60

Qy       181 GAACCAGGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCAGGA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 GluProGlyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGly  80

Qy       241 AATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 AsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLysSer 100

Qy       301 CCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 ProAspGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAla 120

Qy       361 CGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 ArgIleLysLysTrpLeuThrPheSerLeuGlyLysGlnValGlyAsnLysPhePheLeu 140

Qy       421 ACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 ThrAsnGlyGluIleMetThrPheGluLysValLysAlaLeuCysValLysPheGlnAla 160

Qy       481 TCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAGGAG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 SerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsnLeuIleLysGlu 180

Qy       541 GAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThrGly 200

Qy       601 AATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAA 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 AsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGlu 220

Qy       661 GATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCAT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AspCysValLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHis 240

Qy       721 CTGGCCGTCTGTGAGTTCCCTATC 744
             ||||||||||||||||||||||||
Db       241 LeuAlaValCysGluPheProIle 248
```

RESULT 2
US-09-011-735-1
; Sequence 1, Application US/09011735B
; Patent No. 6110708
; GENERAL INFORMATION:
; APPLICANT: Wakamiya, No. 6110708utaka
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof
; FILE REFERENCE: 19036/34548
; CURRENT APPLICATION NUMBER: US/09/011,735B
; CURRENT FILING DATE: 1998-05-22
; EARLIER APPLICATION NUMBER: JP 7-209698
; EARLIER FILING DATE: 1995-08-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Bovine
US-09-011-735-1

Alignment Scores:			
Pred. No.:	1.15e-36	Length:	351
Score:	442.50	Matches:	91
Percent Similarity:	56.77%	Conservative:	39
Best Local Similarity:	39.74%	Mismatches:	74

Query Match:	32.42%	Indels:	25
DB:	3	Gaps:	6

US-10-054-536-2 (1-747) x US-09-011-735-1 (1-351)

```
Qy       118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
             :::||||||||||   ||||||||||||                |||   ||||||||||||
Db       126 AlaProGlyIleGlnGlyPheProGlyPro---------SerGlyLeuLysGlyGluLys 142

Qy       178 GGGGAACCAGGCCAA-----------GGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAG 225
             |||   ||||||:::           |||     |||:::   ||||||   |||
Db       143 GlyAlaProGlyGluThrGlyAlaProGlyArgAlaGlyValThrGlyProSerGlyAla 162

Qy       226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
             :::||||||   |||     |||     |||   ||||||   ||||||   :::|||
Db       163 IleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGly 182

Qy       286 GACCCTGGAAAAAGT---CCGGATGGTGATAGTAGCCTGGCTGCC--------------- 327
             |||||||||::::::     |||:::|||   ||||||
Db       183 AspProGlyGluThrGlyAlaSerGlyGluSerGlyLeuAlaGluValAsnAlaLeuLys 202

Qy       328 ----------------------TCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
                                    :::     |||     ::::::
Db       203 GlnArgValThrIleLeuAspGlyHisLeuArgArgPheGlnAsnAlaPheSerGlnTyr 222

Qy       367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
             ||||||   :::   |||     |||:::   ||||||   |||   |||   |||
Db       223 LysLysAlaValLeuPheProAspGlyGlnAlaValGlyGluLysIlePheLysThrAla 242

Qy       427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
             |||   :::   ::::::       :::   ||||||   :::   :::   :::
Db       243 GlyAlaValLysSerTyrSerAspAlaGluGlnLeuCysArgGluAlaLysGlyGlnLeu 262

Qy       487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG--------- 537
             |||:::||||||::::::||||||||   |||:::     :::::::::
Db       263 AlaSerProArgSerSerAlaGluAsnGluAlaValThrGlnMetValArgAlaGlnGlu 282

Qy       538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
             :::   |||:::|||   :::   |||       |||||||||:::|||         |||
Db       283 LysAsnAlaTyrLeuSerMetAsnAspIleSerThrGluGlyArgPheThrTyrProThr 302

Qy       598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT------GGT 651
             |||     |||   |||:::||||||   :::|||||||||||||||:::     |||
Db       303 GlyGluIleLeuValTyrSerAsnTrpAlaAspGlyGluProAsnAsnSerAspGluGly 322

Qy       652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
                 |||:::||||||   :::     :::|||:::|||||||||||||||||||||
Db       323 GlnProGluAsnCysValGluIleProAspGlyLysTrpAsnAspValProCysSer 342

Qy       712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
                 |||   :::||||||||||
Db       343 LysGlnLeuLeuValIleCysGluPhe 351
```

RESULT 3
US-09-029-156-1
; Sequence 1, Application US/09029156
; Patent No. 6365342
; GENERAL INFORMATION:
; APPLICANT: WAKAMIYA, No. 6365342utaka
; TITLE OF INVENTION: RECOMBINANT CONGLUTININ AND PRODUCING
; TITLE OF INVENTION: METHOD THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/029,156
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/JP96/00173
;       FILING DATE:
;       APPLICATION NUMBER:  PCT/JP95/02035
;       FILING DATE:  02-OCT-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JPA - 209698
;       FILING DATE:  17-AUG-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gass, David A.
;       REGISTRATION NUMBER:  38,153
;       REFERENCE/DOCKET NUMBER:  19036/34546
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312) 474-6300
;       TELEFAX:  (312) 474-0448
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  351 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-029-156-1

Alignment Scores:
Pred. No.:                1.15e-36      Length:          351
Score:                    442.50        Matches:         91
Percent Similarity:       56.77%        Conservative:    39
Best Local Similarity:    39.74%        Mismatches:      74
Query Match:              32.42%        Indels:          25
DB:                       4             Gaps:            6

US-10-054-536-2 (1-747) x US-09-029-156-1 (1-351)

Qy        118 TCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAG 177
              :::|||||||||   ||||||||||||                |||   ||||||||||||
Db        126 AlaProGlyIleGlnGlyPheProGlyPro---------SerGlyLeuLysGlyGluLys 142

Qy        178 GGGGAACCAGGCCAA------------GGGCTCAGAGGCTTACAGGGCCCCCCCTGGAAAG 225
              |||   ||||||:::             |||     |||:::   ||||||   |||
Db        143 GlyAlaProGlyGluThrGlyAlaProGlyArgAlaGlyValThrGlyProSerGlyAla 162

Qy        226 TTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGA 285
              :::||||||   |||      |||      |||   ||||||   ||||||   :::|||
Db        163 IleGlyProGlnGlyProSerGlyAlaArgGlyProProGlyLeuLysGlyAspArgGly 182

Qy        286 GACCCTGGAAAAAGT---CCGGATGGTGATAGTAGCCTGGCTGCC--------------- 327
              |||||||||::::::         |||:::|||   ||||||
Db        183 AspProGlyGluThrGlyAlaSerGlyGluSerGlyLeuAlaGluValAsnAlaLeuLys 202

Qy        328 --------------------TCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC 366
                                  :::     |||         ::::::
Db        203 GlnArgValThrIleLeuAspGlyHisLeuArgArgPheGlnAsnAlaPheSerGlnTyr 222

Qy        367 AAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              ||||||   :::   |||      |||:::   ||||||   |||   |||   |||
Db        223 LysLysAlaValLeuPheProAspGlyGlnAlaValGlyGluLysIlePheLysThrAla 242

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||   :::   ::::::         :::   ||||||   :::   :::      :::
Db        243 GlyAlaValLysSerTyrSerAspAlaGluGlnLeuCysArgGluAlaLysGlyGlnLeu 262

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG--------- 537
              |||:::||||||::::::|||||||||   |||:::     ::::::::
Db        263 AlaSerProArgSerSerAlaGluAsnGluAlaValThrGlnMetValArgAlaGlnGlu 282

Qy        538 GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACA 597
              :::   |||:::|||   :::   |||      |||||||||:::|||         |||

Db        283 LysAsnAlaTyrLeuSerMetAsnAspIleSerThrGluGlyArgPheThrTyrProThr 302

Qy        598 GGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT------GGT 651
              |||      |||   |||:::||||||   :::|||||||||||||||:::      |||
Db        303 GlyGluIleLeuValTyrSerAsnTrpAlaAspGlyGluProAsnAsnSerAspGluGly 322

Qy        652 TCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCC 711
                    |||:::||||||   :::      :::|||:::|||||||||||||||||||||
Db        323 GlnProGluAsnCysValGluIlePheProAspGlyLysTrpAsnAspValProCysSer 342

Qy        712 ACCTCCCATCTGGCCGTCTGTGAGTTC 738
                    |||   :::||||||||||
Db        343 LysGlnLeuLeuValIleCysGluPhe 351

RESULT 4
5514582-16
;Patent No. 5514582
;     APPLICANT: CAPON, DANIEL J.;LASKY, LAURENCE A.
;     TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
;     NUMBER OF SEQUENCES: 43
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/185,670
;       FILING DATE: 21-JAN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 986,931
;       FILING DATE: 08-DEC-1992
;       APPLICATION NUMBER: 808,122
;       FILING DATE: 16-DEC-1991
;       APPLICATION NUMBER: 440,625
;       FILING DATE: 22-NOV-1989
;       APPLICATION NUMBER: 315,015
;       FILING DATE: 23-FEB-1989
;SEQ ID NO:16:
;       LENGTH: 108
5514582-16

Alignment Scores:
Pred. No.:                2.35e-28      Length:          108
Score:                    359.00        Matches:         66
Percent Similarity:       74.07%        Conservative:    14
Best Local Similarity:    61.11%        Mismatches:      28
Query Match:              26.30%        Indels:          0
DB:                       6             Gaps:            0

US-10-054-536-2 (1-747) x 5514582-16 (1-108)

Qy        409 AAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTC 468
              |||||||||:::||||||   |||   |||   |||   ||||||||||||||||||
Db          1 LysPhePheValThrAsnHisGluArgMetProPheSerLysValLysAlaLeuCysSer 20

Qy        469 AAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAAT 528
              :::   :::   :::||||||   |||:::||||||   ||||||   |||||||||
Db         21 GluLeuArgGlyThrValAlaIleProLysAsnAlaGluGluAsnLysAlaIleGlnGlu 40

Qy        529 CTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTG 588
              :::   |||      ||||||||||||||||||||||||   ||||||||||||||:::
Db         41 ValAlaLysThrSerAlaPheLeuGlyIleThrAspGluValThrGluGlyGlnPheMet 60

Qy        589 GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT 648
                 :::||||||   ||||||||||||:::|||||||   :::   |||||||||:::
Db         61 TyrValThrGlyGlyArgLeuThrTyrSerAsnTrpLysLysAspGluProAsnAspHis 80

Qy        649 GGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGC 708
              ||||||   ||||||||||||   ::::::   ||||||   |||||||||:::   |||
Db         81 GlySerGlyGluAspCysValThrIleValAspAsnGlyLeuTrpAsnAspIleSerCys 100

Qy        709 TCCACCTCCCATCTGGCCGTCTGT 732
                    ||||||   |||||||||
Db        101 GlnAlaSerHisThrAlaValCys 108

RESULT 5
5514582-17
```

;Patent No. 5514582
; APPLICANT: CAPON, DANIEL J.;LASKY, LAURENCE A.
; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 43
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,670
; FILING DATE: 21-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 986,931
; FILING DATE: 08-DEC-1992
; APPLICATION NUMBER: 808,122
; FILING DATE: 16-DEC-1991
; APPLICATION NUMBER: 440,625
; FILING DATE: 22-NOV-1989
; APPLICATION NUMBER: 315,015
; FILING DATE: 23-FEB-1989
;SEQ ID NO:17:
; LENGTH: 107
5514582-17

Alignment Scores:

Pred. No.:	3.49e-27	Length:	107
Score:	347.50	Matches:	66
Percent Similarity:	73.15%	Conservative:	13
Best Local Similarity:	61.11%	Mismatches:	28
Query Match:	25.46%	Indels:	1
DB:	6	Gaps:	1

US-10-054-536-2 (1-747) x 5514582-17 (1-107)

```
Qy        409 AAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTC 468
              |||||||||:::||||||   |||   |||   |||   ||||||||||||||||||
Db          1 LysPhePheValThrAsnHisGluArgMetProPheSerLysValLysAlaLeuCysSer 20

Qy        469 AAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAAT 528
              :::   :::   :::||||||   ||||||||||   ||||||   ||||||||
Db         21 GluLeuArgGlyThrValAlaIleProArgAsnAlaGluGluAsnLysAlaIleGlnGlu 40

Qy        529 CTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTG 588
              :::   |||      ||||||||||||||||||||||   |||||||||||||||:::
Db         41 ValAlaLysThrSerAlaPheLeuGlyIleThrAspGluValThrGluGlyGlnPheMet 60

Qy        589 GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT 648
              :::||||||   ||||||||||:::||||||   :::   |||||||||:::
Db         61 TyrValThrGlyGlyArgLeuThrTyrSerAsnTrpLysLysAspGluProAsnAspHis 80

Qy        649 GGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGC 708
              |||      ||||||||||   ::::::   ||||||   ||||||||:::   |||
Db         81 Gly---HisGluAspCysValThrIleValAspAsnGlyLeuTrpAsnAspIleSerCys 99

Qy        709 TCCACCTCCCATCTGGCCGTCTGT 732
              ||||||   |||||||||
Db        100 GlnAlaSerHisThrAlaValCys 107
```

RESULT 6
US-09-011-735-6
; Sequence 6, Application US/09011735B
; Patent No. 6110708
; GENERAL INFORMATION:
; APPLICANT: Wakamiya, No. 6110708utaka
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof
; FILE REFERENCE: 19036/34548
; CURRENT APPLICATION NUMBER: US/09/011,735B
; CURRENT FILING DATE: 1998-05-22
; EARLIER APPLICATION NUMBER: JP 7-209698
; EARLIER FILING DATE: 1995-08-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (2)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (5)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (6)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
US-09-011-735-6

Alignment Scores:

Pred. No.:	6.86e-21	Length:	161
Score:	286.50	Matches:	54
Percent Similarity:	61.15%	Conservative:	31
Best Local Similarity:	38.85%	Mismatches:	49
Query Match:	20.99%	Indels:	5
DB:	3	Gaps:	2

US-10-054-536-2 (1-747) x US-09-011-735-6 (1-161)

```
Qy        337 AAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAA 396
              :::      |||          ::::::   ||||||   :::   |||      |||:::
Db         23 ArgArgPheGlnAsnAlaPheSerGlnTyrLysLysAlaValLeuPheProAspGlyGln 42

Qy        397 CAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAG 456
                 ||||||   |||   |||   |||   |||   :::   ::::::         :::
Db         43 AlaValGlyGluLysIlePheLysThrAlaGlyAlaValLysSerTyrSerAspAlaGlu 62

Qy        457 GCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGA 516
                 ||||||   :::   :::      :::|||:::|||||:::::|||||||||
Db         63 GlnLeuCysArgGluAlaLysGlyGlnLeuAlaSerProArgSerSerAlaGluAsnGlu 82

Qy        517 GCCATTCAGAATCTCATCAAG---------GAGGAAGCCTTCCTGGGCATCACTGATGAG 567
              |||:::      :::::::::         :::   |||:::|||   :::   |||
Db         83 AlaValThrGlnMetValArgAlaGlnGluLysAsnAlaTyrLeuSerMetAsnAspIle 102

Qy        568 AAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAAC 627
                 |||||||||:::|||         |||||      |||   |||:::||||||
Db        103 SerThrGluGlyArgPheThrTyrProThrGlyGluIleLeuValTyrSerAsnTrpAla 122

Qy        628 GAGGGTGAACCCAACAATGCT------GGTTCTGATGAAGATTGTGTATTGCTACTGAAA 681
              :::|||||||||||||||:::      |||      |||:::||||||   :::
Db        123 AspGlyGluProAsnAsnSerAspGluGlyGlnProGluAsnCysValGluIlePhePro 142

Qy        682 AATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTC 738
              :::|||:::|||||||||||||||||||||||        |||   :::|||||||||
Db        143 AspGlyLysTrpAsnAspValProCysSerLysGlnLeuLeuValIleCysGluPhe 161
```

RESULT 7
5514582-20
;Patent No. 5514582
; APPLICANT: CAPON, DANIEL J.;LASKY, LAURENCE A.
; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 43
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,670
; FILING DATE: 21-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 986,931
; FILING DATE: 08-DEC-1992
; APPLICATION NUMBER: 808,122
; FILING DATE: 16-DEC-1991
; APPLICATION NUMBER: 440,625
; FILING DATE: 22-NOV-1989

```
;      APPLICATION NUMBER: 315,015
;      FILING DATE: 23-FEB-1989
;SEQ ID NO:20:
;      LENGTH: 111
5514582-20

Alignment Scores:
Pred. No.:              2.08e-13      Length:          111
Score:                  212.50        Matches:         42
Percent Similarity:     56.25%        Conservative:    21
Best Local Similarity:  37.50%        Mismatches:      44
Query Match:            15.57%        Indels:          5
DB:                     6             Gaps:            3

US-10-054-536-2 (1-747) x 5514582-20 (1-111)

Qy        409 AAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTC 468
              |||   |||   |||||||||:::   :::   |||:::   ::::::   :::|||
Db          1 LysIlePheSerThrAsnGlyGlnSerValAsnPheAspAlaIleArgGluValCysAla 20

Qy        469 AAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAAT---GGAGCCATTCAG 525
              :::           :::|||   ||||||:::      ||||||      ||||||
Db         21 ArgAlaGlyGlyArgIleAlaValProArgSerLeuGluGluAsnGluAlaAlaIleAla 40

Qy        526 AATCTCATCAAGGAGGAA---------GCCTTCCTGGGCATCACTGATGAGAAGACAGAA 576
              :::::::::|||||           |||:::|||||:::   :::     |||
Db         41 SerIleValLysGluArgAsnThrTyrAlaTyrLeuGlyLeuAlaGluGlyProThrAla 60

Qy        577 GGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAA 636
              |||   |||       |||   |||:::   :::   ||||||||||||      ||||||
Db         61 GlyAspPheTyrTyrLeuAspGlyAspProValAsnTyrThrAsnTrpTyrProGlyGlu 80

Qy        637 CCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAAT 696
              |||         |||       |||   ||||||   :::      :::|||:::||||||
Db         81 ProArgGlyGlnGlyArg---GluLysCysValGluMetTyrThrAspGlyLysTrpAsn 99

Qy        697 GACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGT 732
              |||       |||             |||   :::|||
Db        100 AspLysAsnCysLeuGlnTyrArgLeuValIleCys 111

RESULT 8
US-09-602-877A-99
; Sequence 99, Application US/09602877A
; Patent No. 6432707
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF BREAST CANCER
;  FILE REFERENCE:  210121.446C5
;  CURRENT APPLICATION NUMBER: US/09/602,877A
;  CURRENT FILING DATE:  2000-06-22
;  NUMBER OF SEQ ID NOS:  107
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 99
;   LENGTH: 197
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-602-877A-99

Alignment Scores:
Pred. No.:              1.93e-12      Length:          197
Score:                  204.00        Matches:         54
Percent Similarity:     46.63%        Conservative:    29
Best Local Similarity:  30.34%        Mismatches:      65
Query Match:            14.95%        Indels:          30
DB:                     4             Gaps:            5

US-10-054-536-2 (1-747) x US-09-602-877A-99 (1-197)

Qy        271 AAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTG--------- 321
              |||   :::|||           :::      |||   |||      |||

Db         28 LysAlaArgLysHisSerLysArgArgValArgAspLysAspGlyAspLeuLysThrGln 47

Qy        322 ----------------------------GCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAA 354
                                          |||      |||   :::|||||||||||||
Db         48 IleGluLysLeuTrpThrGluValAsnAlaLeuLysGluIleGlnAlaLeuGlnThrVal 67

Qy        355 ATGGCACGTATCAAAAAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTC 414
                 |||                                |||   :::|||      |||
Db         68 CysLeuArg----------------------------GlyThrLysValHisLysLysCys 78

Qy        415 TTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTC 474
              :::|||   :::      :::      |||   :::          |||:::
Db         79 TyrLeuAlaSerGluGlyLeuLysHisPheHisGluAlaAsnGluAspCysIleSerLys 98

Qy        475 CAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC 534
                       :::      |||||||||:::   |||      |||:::|||:::
Db         99 GlyGlyIleLeuValIleProArgAsnSerAspGluIleAsnAlaLeuGlnAspTyrGly 118

Qy        535 AAG-----------------GAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAA 576
              |||                    :::   :::|||||||||   |||      ||||||
Db        119 LysArgSerLeuProGlyValAsnAspPheTrpLeuGlyIleAsnAspMetValThrGlu 138

Qy        577 GGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAA 636
              |||:::|||||||||:::   |||      :::::::::   ||||||:::      :::
Db        139 GlyLysPheValAspValAsnGlyIleAlaIleSerPheLeuAsnTrpAspArgAlaGln 158

Qy        637 CCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAAT------GGCCAG 690
              |||   |||   |||       |||:::|||||||||      ::::::      |||:::
Db        159 Pro---AsnGlyGlyLysArgGluAsnCysValLeuPheSerGlnSerAlaGlnGlyLys 177

Qy        691 TGGAATGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATC 744
              |||:::|||       |||   :::|||          :::|||||||||   |||
Db        178 TrpSerAspGluAlaCysArgSerSerLysArgTyrIleCysGluPheThrIle 195

RESULT 9
US-08-365-103B-4
; Sequence 4, Application US/08365103B
; Patent No. 5766943
;  GENERAL INFORMATION:
;    APPLICANT:  Lynch, Richard G
;    APPLICANT:  Nunez, Raphael D.
;    APPLICANT:  Yodoi, Jungi
;    TITLE OF INVENTION:  DNA Sequences for Soluble Froms of CD23
;    TITLE OF INVENTION:  and Methods of Use for Same
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zarley, McKee, Thomte, Voorhees & Sease
;      STREET:  801 Grand Ave. Suite 3200
;      CITY:  Des Moines
;      STATE:  Iowa
;      COUNTRY:  United States
;      ZIP:  50309
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,103B
;      FILING DATE:  28-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nebel, Heidi S.
;      REGISTRATION NUMBER:  37,719
;      REFERENCE/DOCKET NUMBER:  Uirf N5-24
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (515) 288-3667
;      TELEFAX:  (515) 288-1338
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  287 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-365-103B-4

Alignment Scores:
Pred. No.:                4.04e-12        Length:          287
Score:                    201.50          Matches:         51
Percent Similarity:       46.59%          Conservative:    31
Best Local Similarity:    28.98%          Mismatches:      47
Query Match:              14.76%          Indels:          47
DB:                       1               Gaps:            7

US-10-054-536-2 (1-747) x US-08-365-103B-4 (1-287)

Qy         316 AGCCTGGCTGCCTCAGAAAGAAAAGCT----------------------CTGCAAACAGAA 354
               ||||||      :::|||::::::                        ||||||   |||
Db         108 SerLeuGlyLeuAsnGluLysArgThrAlaSerAspSerLeuGluLysLeuGlnGluGlu 127

Qy         355 ATGGCACGTATC---------------------------------------------- 366
               :::|||::::::
Db         128 ValAlaLysLeuTrpIleGluIleLeuIleSerLysGlyThrAlaCysAsnIleCysPro 147

Qy         367 AAAAAGTGGCTGACCTTCTCT---------------CTGGGCAAACAAGTTGGGAACAAG 411
               |||   ||||||   |||                 ||||||      |||:::|||
Db         148 LysAsnTrpLeuHisPheGlnGlnLysCysTyrTyrPheGlyLys------GlySerLys 165

Qy         412 TTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAG 471
                 ::::::                          :::   :::      |||
Db         166 GlnTrpIle-------------------------GlnAlaArgPheAlaCysSerAsp 176

Qy         472 TTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGA---GCCATTCAGAAT 528
               |||      :::   :::      :::      |||           :::|||:::
Db         177 LeuGlnGlyArgLeuValSerIleHisSerGlnLysGluGlnAspPheLeuMetGlnHis 196

Qy         529 CTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTG 588
               :::   |||:::::::::::::::|||:::   |||         ||||||:::||||||
Db         197 IleAsnLysLysAspSerTrpIleGlyLeuGlnAspLeuAsnMetGluGlyGluPheVal 216

Qy         589 GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT 648
                  |||:::   :::   |||:::|||||||||   ||||||||||||||
Db         217 TrpSerAspGlySerProValGlyTyrSerAsnTrpAsnProGlyGluProAsnAsnGly 236

Qy         649 GGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGC 708
               |||      ||||||||||||:::::      :::||||||||||||||      |||
Db         237 GlyGlnGlyGluAspCysValMetMetArgGlySerGlyGlnTrpAsnAspAlaPheCys 256

Qy         709 TCCACC--------------------TCCCATCTGGCCGTCTGTGAG 735
                  :::                   ||||||   ||||||
Db         257 ArgSerTyrLeuAspAlaTrpValCysGluGlnLeuAlaThrCysGlu 272

RESULT 10
US-08-365-103B-6
; Sequence 6, Application US/08365103B
; Patent No. 5766943
;  GENERAL INFORMATION:
;    APPLICANT:  Lynch, Richard G
;    APPLICANT:  Nunez, Raphael D.
;    APPLICANT:  Yodoi, Jungi
;    TITLE OF INVENTION:  DNA Sequences for Soluble Froms of CD23
;    TITLE OF INVENTION:  and Methods of Use for Same
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zarley, McKee, Thomte, Voorhees & Sease
;      STREET:  801 Grand Ave. Suite 3200
;      CITY:  Des Moines
;      STATE:  Iowa
;      COUNTRY:  United States
;      ZIP:  50309
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,103B
;      FILING DATE:  28-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nebel, Heidi S.
;      REGISTRATION NUMBER:  37,719
;      REFERENCE/DOCKET NUMBER:  Uirf N5-24
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (515) 288-3667
;      TELEFAX:  (515) 288-1338
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  300 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-365-103B-6

Alignment Scores:
Pred. No.:                4.12e-12        Length:          300
Score:                    201.50          Matches:         51
Percent Similarity:       46.59%          Conservative:    31
Best Local Similarity:    28.98%          Mismatches:      47
Query Match:              14.76%          Indels:          47
DB:                       1               Gaps:            7

US-10-054-536-2 (1-747) x US-08-365-103B-6 (1-300)

Qy         316 AGCCTGGCTGCCTCAGAAAGAAAAGCT----------------------CTGCAAACAGAA 354
               ||||||      :::|||::::::                        ||||||   |||
Db         121 SerLeuGlyLeuAsnGluLysArgThrAlaSerAspSerLeuGluLysLeuGlnGluGlu 140

Qy         355 ATGGCACGTATC---------------------------------------------- 366
               :::|||::::::
Db         141 ValAlaLysLeuTrpIleGluIleLeuIleSerLysGlyThrAlaCysAsnIleCysPro 160

Qy         367 AAAAAGTGGCTGACCTTCTCT---------------CTGGGCAAACAAGTTGGGAACAAG 411
               |||   ||||||   |||                 ||||||      |||:::|||
Db         161 LysAsnTrpLeuHisPheGlnGlnLysCysTyrTyrPheGlyLys------GlySerLys 178

Qy         412 TTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAG 471
                 ::::::                          :::   :::      |||
Db         179 GlnTrpIle-------------------------GlnAlaArgPheAlaCysSerAsp 189

Qy         472 TTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGA---GCCATTCAGAAT 528
               |||      :::   :::      :::      |||           :::|||:::
Db         190 LeuGlnGlyArgLeuValSerIleHisSerGlnLysGluGlnAspPheLeuMetGlnHis 209

Qy         529 CTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTG 588
               :::   |||:::::::::::::::|||:::   |||         ||||||:::||||||
Db         210 IleAsnLysLysAspSerTrpIleGlyLeuGlnAspLeuAsnMetGluGlyGluPheVal 229

Qy         589 GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT 648
                  |||:::   :::   |||:::|||||||||   ||||||||||||||
Db         230 TrpSerAspGlySerProValGlyTyrSerAsnTrpAsnProGlyGluProAsnAsnGly 249

Qy         649 GGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGC 708
               |||      ||||||||||||:::::      :::||||||||||||||      |||
Db         250 GlyGlnGlyGluAspCysValMetMetArgGlySerGlyGlnTrpAsnAspAlaPheCys 269

Qy         709 TCCACC--------------------TCCCATCTGGCCGTCTGTGAG 735
                  :::                   ||||||   ||||||
Db         270 ArgSerTyrLeuAspAlaTrpValCysGluGlnLeuAlaThrCysGlu 285

RESULT 11
US-08-365-103B-2
; Sequence 2, Application US/08365103B
; Patent No. 5766943
;  GENERAL INFORMATION:
;    APPLICANT:  Lynch, Richard G
;    APPLICANT:  Nunez, Raphael D.
;    APPLICANT:  Yodoi, Jungi
;    TITLE OF INVENTION:  DNA Sequences for Soluble Froms of CD23
```

; TITLE OF INVENTION: and Methods of Use for Same
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease
; STREET: 801 Grand Ave. Suite 3200
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: United States
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,103B
; FILING DATE: 28-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Nebel, Heidi S.
; REGISTRATION NUMBER: 37,719
; REFERENCE/DOCKET NUMBER: Uirf N5-24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 288-3667
; TELEFAX: (515) 288-1338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-365-103B-2

Alignment Scores:

Pred. No.:	4.26e-12	Length:	327
Score:	201.50	Matches:	51
Percent Similarity:	46.59%	Conservative:	31
Best Local Similarity:	28.98%	Mismatches:	47
Query Match:	14.76%	Indels:	47
DB:	1	Gaps:	7

US-10-054-536-2 (1-747) x US-08-365-103B-2 (1-327)

```
Qy        316 AGCCTGGCTGCCTCAGAAAGAAAAGCT--------------------CTGCAAACAGAA 354
              ||||||      :::|||::::::                      ||||||   |||
Db        148 SerLeuGlyLeuAsnGluLysArgThrAlaSerAspSerLeuGluLysLeuGlnGluGlu 167

Qy        355 ATGGCACGTATC------------------------------------------------ 366
              :::|||::::::
Db        168 ValAlaLysLeuTrpIleGluIleLeuIleSerLysGlyThrAlaCysAsnIleCysPro 187

Qy        367 AAAAAGTGGCTGACCTTCTCT---------------CTGGGCAAACAAGTTGGGAACAAG 411
              |||   ||||||   |||                 ||||||     |||:::|||
Db        188 LysAsnTrpLeuHisPheGlnGlnLysCysTyrTyrPheGlyLys------GlySerLys 205

Qy        412 TTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAG 471
                 ::::::                          :::  :::     |||
Db        206 GlnTrpIle-------------------------GlnAlaArgPheAlaCysSerAsp 216

Qy        472 TTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGGA---GCCATTCAGAAT 528
                 |||      :::   :::      :::      |||            :::|||:::
Db        217 LeuGlnGlyArgLeuValSerIleHisSerGlnLysGluGlnAspPheLeuMetGlnHis 236

Qy        529 CTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTG 588
              :::   |||:::::::::::::::|||:::   |||         |||||||:::|||||
Db        237 IleAsnLysLysAspSerTrpIleGlyLeuGlnAspLeuAsnMetGluGlyGluPheVal 256

Qy        589 GATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCT 648
                  |||:::   :::   |||:::|||||||||   ||||||||||||||
Db        257 TrpSerAspGlySerProValGlyTyrSerAsnTrpAsnProGlyGluProAsnAsnGly 276

Qy        649 GGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGC 708
              |||      ||||||||||||::::::      :::||||||||||||||||      |||
```

```
Db        277 GlyGlnGlyGluAspCysValMetMetArgGlySerGlyGlnTrpAsnAspAlaPheCys 296

Qy        709 TCCACC--------------------TCCCATCTGGCCGTCTGTGAG 735
                  :::                   ||||||   ||||||
Db        297 ArgSerTyrLeuAspAlaTrpValCysGluGlnLeuAlaThrCysGlu 312
```

RESULT 12
US-09-535-521-20
; Sequence 20, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
; APPLICANT: Weber, Eric R.
; APPLICANT: McCall, Catherine A.
; TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; FILE REFERENCE: AL-5
; CURRENT APPLICATION NUMBER: US/09/535,521
; CURRENT FILING DATE: 2000-03-24
; EARLIER APPLICATION NUMBER: 60/125,913
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-535-521-20

Alignment Scores:

Pred. No.:	1.15e-11	Length:	208
Score:	196.50	Matches:	67
Percent Similarity:	44.35%	Conservative:	35
Best Local Similarity:	29.13%	Mismatches:	77
Query Match:	14.40%	Indels:	52
DB:	4	Gaps:	11

US-10-054-536-2 (1-747) x US-09-535-521-20 (1-208)

```
Qy        122 CAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGG 181
              ||||||:::      ||||||           :::   |||      |||::::::
Db          4 GlnAlaAlaGlnValSerGlnAspMetLysGluIleGlnAla------GluGlnLysArg 21

Qy        182 AACCAGGCCAAGGGCTCAGAG-----------------GCTTACAGGGCCCCC------ 217
              :::|||:::   ||||||                  |||   |||:::
Db         22 MetLysAlaGlnAspSerGluLeuSerGlnAsnLeuAspAlaLeuArgSerAspLeuAsn 41

Qy        218 ---CTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGG 274
                 |||:::|||      |||   |||          |||   |||      :::|||
Db         42 AsnLeuLysSerGlnSerLeuAsnGluArgSerThrAlaLeu---HisSerLeuGluArg 60

Qy        275 GCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAA 334
                 ::::::|||   :::||||||:::                  |||:::
Db         61 LeuGlnGluGlu---ValGluLysLeu-----------------TrpMet--------- 70

Qy        335 GAAAAGCTCTGCAAACAGAAATGGCACGTATCAAA----------------------- 369
                           :::   |||||||||
Db         71 -----------------GluLeuHisValSerAsnGlySerGluCysAsnThrCysPro 84

Qy        370 ---AAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
                 |||||||||   |||                  |||   :::
Db         85 GluLysTrpLeuAsnPheGln-----------------ArgLysCysTyrTyrPheGly 98

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
                 |||      :::   :::      :::      |||   |||   |||      :::
Db         99 GluGluProLysLysTrpIleGlnAlaArgPheAlaCysSerLysLeuGlnGlyArgLeu 118

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---AAGGAGGAA 543
                 |||:::      :::      |||      :::                |||:::
Db        119 AlaSerIleHisSerGlnGluGluGlnAspPheLeuAlaArgTyrAlaAsnLysLysGly 138

Qy        544 GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAAT 603
                 ::::::|||:::   |||         ||||||:::|||:::   :::      |||
Db        139 ThrTrpIleGlyLeuArgAspLeuAspArgGluGlyGluPheIleTrpMetAspGluAsn 158
```

```
Qy        604 AGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGAT 663
              |||   |||:::||||||      ||||||||||||||   |||      ||||||
Db        159 ProLeuAsnTyrSerAsnTrpArgProGlyGluProAsnAsnGlyGlyGlnGlyGluAsp 178

Qy        664 TGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC------ 717
              ||||||::::::      :::|||||||||||||||      |||   :::|||
Db        179 CysValMetMetGlnGlySerGlyGlnTrpAsnAspAlaPheCysGlySerSerLeuAsp 198

Qy        718 ---------------CATCTGGCCGTCTGT 732
                             ||||||   |||
Db        199 GlyTrpValCysAspArgLeuAlaThrCys 208

RESULT 13
US-09-535-521-2
; Sequence 2, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 292
;   TYPE: PRT
;   ORGANISM: Canis familiaris
US-09-535-521-2

Alignment Scores:
Pred. No.:                1.32e-11      Length:          292
Score:                    196.50        Matches:         67
Percent Similarity:       44.35%        Conservative:    35
Best Local Similarity:    29.13%        Mismatches:      77
Query Match:              14.40%        Indels:          52
DB:                       4             Gaps:            11

US-10-054-536-2 (1-747) x US-09-535-521-2 (1-292)

Qy        122 CAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGG 181
              ||||||:::      ||||||           :::   |||      |||::::::
Db         88 GlnAlaAlaGlnValSerGlnAspMetLysGluIleGlnAla------GluGlnLysArg 105

Qy        182 AACCAGGCCAAGGGCTCAGAG-----------------GCTTACAGGGCCCCC------ 217
              :::|||:::   ||||||                  |||   |||:::
Db        106 MetLysAlaGlnAspSerGluLeuSerGlnAsnLeuAspAlaLeuArgSerAspLeuAsn 125

Qy        218 ---CTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGG 274
                 |||:::|||      |||   |||           |||   |||      :::|||
Db        126 AsnLeuLysSerGlnSerLeuAsnGluArgSerThrAlaLeu---HisSerLeuGluArg 144

Qy        275 GCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAA 334
              ::::::|||   :::||||||:::                  |||:::
Db        145 LeuGlnGluGlu---ValGluLysLeu-----------------TrpMet--------- 154

Qy        335 GAAAAGCTCTGCAAACAGAAATGGCACGTATCAAA----------------------- 369
                          :::   |||||||||
Db        155 -----------------GluLeuHisValSerAsnGlySerGluCysAsnThrCysPro 168

Qy        370 ---AAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
              |||||||||   |||                  |||   :::
Db        169 GluLysTrpLeuAsnPheGln-----------------ArgLysCysTyrTyrPheGly 182

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||         :::   :::   :::      |||   |||   |||      :::
Db        183 GluGluProLysLysTrpIleGlnAlaArgPheAlaCysSerLysLeuGlnGlyArgLeu 202

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---AAGGAGGAA 543
              |||:::      :::    |||      :::           |||:::
Db        203 AlaSerIleHisSerGlnGluGluGlnAspPheLeuAlaArgTyrAlaAsnLysLysGly 222

Qy        544 GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAAT 603
                 ::::::|||:::   |||         ||||||:::|||:::   :::      |||
Db        223 ThrTrpIleGlyLeuArgAspLeuAspArgGluGlyGluPheIleTrpMetAspGluAsn 242

Qy        604 AGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGAT 663
              |||   |||:::||||||      ||||||||||||||   |||      ||||||
Db        243 ProLeuAsnTyrSerAsnTrpArgProGlyGluProAsnAsnGlyGlyGlnGlyGluAsp 262

Qy        664 TGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC------ 717
              ||||||::::::      :::|||||||||||||||      |||   :::|||
Db        263 CysValMetMetGlnGlySerGlyGlnTrpAsnAspAlaPheCysGlySerSerLeuAsp 282

Qy        718 ---------------CATCTGGCCGTCTGT 732
                             ||||||   |||
Db        283 GlyTrpValCysAspArgLeuAlaThrCys 292

RESULT 14
US-09-535-521-5
; Sequence 5, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 292
;   TYPE: PRT
;   ORGANISM: Canis familiaris
US-09-535-521-5

Alignment Scores:
Pred. No.:                1.32e-11      Length:          292
Score:                    196.50        Matches:         67
Percent Similarity:       44.35%        Conservative:    35
Best Local Similarity:    29.13%        Mismatches:      77
Query Match:              14.40%        Indels:          52
DB:                       4             Gaps:            11

US-10-054-536-2 (1-747) x US-09-535-521-5 (1-292)

Qy        122 CAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGCACCAAGGGAGAAAAGGGGG 181
              ||||||:::      ||||||           :::   |||      |||::::::
Db         88 GlnAlaAlaGlnValSerGlnAspMetLysGluIleGlnAla------GluGlnLysArg 105

Qy        182 AACCAGGCCAAGGGCTCAGAG-----------------GCTTACAGGGCCCCC------ 217
              :::|||:::   ||||||                  |||   |||:::
Db        106 MetLysAlaGlnAspSerGluLeuSerGlnAsnLeuAspAlaLeuArgSerAspLeuAsn 125

Qy        218 ---CTGGAAAGTTGGGGCCTCCAGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGG 274
                 |||:::|||      |||   |||           |||   |||      :::|||
Db        126 AsnLeuLysSerGlnSerLeuAsnGluArgSerThrAlaLeu---HisSerLeuGluArg 144

Qy        275 GCCAAAAAGGAGACCCTGGAAAAAGTCCGGATGGTGATAGTAGCCTGGCTGCCTCAGAAA 334
              ::::::|||   :::||||||:::                  |||:::
Db        145 LeuGlnGluGlu---ValGluLysLeu-----------------TrpMet--------- 154

Qy        335 GAAAAGCTCTGCAAACAGAAATGGCACGTATCAAA----------------------- 369
                          :::   |||||||||
Db        155 -----------------GluLeuHisValSerAsnGlySerGluCysAsnThrCysPro 168

Qy        370 ---AAGTGGCTGACCTTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAAT 426
```

```
                ||||||||   |||                  |||   :::
Db        169 GluLysTrpLeuAsnPheGln-----------------ArgLysCysTyrTyrPheGly 182

Qy        427 GGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTG 486
              |||       :::   :::   :::      |||   |||   |||      :::
Db        183 GluGluProLysLysTrpIleGlnAlaArgPheAlaCysSerLysLeuGlnGlyArgLeu 202

Qy        487 GCCACCCCCAGGAATGCTGCAGAGAATGGAGCCATTCAGAATCTCATC---AAGGAGGAA 543
              |||:::      :::      |||         :::                |||:::
Db        203 AlaSerIleHisSerGlnGluGluGlnAspPheLeuAlaArgTyrAlaAsnLysLysGly 222

Qy        544 GCCTTCCTGGGCATCACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAAT 603
              ::::::|||:::   |||       ||||||:::|||:::   :::      |||
Db        223 ThrTrpIleGlyLeuArgAspLeuAspArgGluGlyGluPheIleTrpMetAspGluAsn 242

Qy        604 AGACTGACCTACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGAT 663
              |||   |||:::|||||      |||||||||||||||   |||      ||||||
Db        243 ProLeuAsnTyrSerAsnTrpArgProGlyGluProAsnAsnGlyGlyGlnGlyGluAsp 262

Qy        664 TGTGTATTGCTACTGAAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC------ 717
              ||||||::::::      :::|||||||||||||||      |||   :::|||
Db        263 CysValMetMetGlnGlySerGlyGlnTrpAsnAspAlaPheCysGlySerSerLeuAsp 282

Qy        718 ---------------CATCTGGCCGTCTGT 732
                             ||||||   |||
Db        283 GlyTrpValCysAspArgLeuAlaThrCys 292

RESULT 15
US-09-535-521-17
; Sequence 17, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
;  APPLICANT: Weber, Eric R.
;  APPLICANT: McCall, Catherine A.
;  TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IgE RECEPTOR (CANINE CD23)
;  TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: AL-5
;  CURRENT APPLICATION NUMBER: US/09/535,521
;  CURRENT FILING DATE: 2000-03-24
;  EARLIER APPLICATION NUMBER: 60/125,913
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
;   LENGTH: 187
;   TYPE: PRT
;   ORGANISM: Canis familiaris
US-09-535-521-17

Alignment Scores:
Pred. No.:                1.24e-11        Length:          187
Score:                    196.00          Matches:         47
Percent Similarity:       44.24%          Conservative:    26
Best Local Similarity:    28.48%          Mismatches:      62
Query Match:              14.36%          Indels:          30
DB:                       4               Gaps:            4

US-10-054-536-2 (1-747) x US-09-535-521-17 (1-187)

Qy        310 GATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATC--- 366
              :::   |||   |||            :::   ||||||   |||:::   ::::::
Db         29 GluArgSerThrAlaLeuHisSerLeuGluArgLeuGlnGluGluValGluLysLeuTrp 48

Qy        367 ---------------------------------------------AAAAAGTGGCTGACC 381
                                                          :::|||||||||
Db         49 MetGluLeuHisValSerAsnGlySerGluCysAsnThrCysProGluLysTrpLeuAsn 68

Qy        382 TTCTCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACC 441
              |||                        |||   :::            |||
Db         69 PheGln-----------------ArgLysCysTyrTyrPheGlyGluGluProLysLys 82

Qy        442 TTTGAAAAAGTGAAGGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAAT 501
              :::   :::   :::      |||   |||   |||      :::|||:::      :::

Db         83 TrpIleGlnAlaArgPheAlaCysSerLysLeuGlnGlyArgLeuAlaSerIleHisSer 102

Qy        502 GCTGCAGAGAATGGAGCCATTCAGAATCTCATC---AAGGAGGAAGCCTTCCTGGGCATC 558
              |||          :::                |||:::      ::::::|||:::
Db        103 GlnGluGluGlnAspPheLeuAlaArgTyrAlaAsnLysLysGlyThrTrpIleGlyLeu 122

Qy        559 ACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACA 618
              |||         ||||||:::|||:::   :::      |||   |||   |||:::
Db        123 ArgAspLeuAspArgGluGlyGluPheIleTrpMetAspGluAsnProLeuAsnTyrSer 142

Qy        619 AACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTG 678
              ||||||      ||||||||||||||   |||      ||||||||||:::::::
Db        143 AsnTrpArgProGlyGluProAsnAsnGlyGlyGlnGlyGluAspCysValMetMetGln 162

Qy        679 AAAAATGGCCAGTGGAATGACGTCCCCTGCTCCACCTCC---------------------- 717
              :::|||||||||||||||      |||   :::|||
Db        163 GlySerGlyGlnTrpAsnAspAlaPheCysGlySerSerLeuAspGlyTrpValCysAsp 182

Qy        718 CATCTGGCCGTCTGT 732
              ||||||   |||
Db        183 ArgLeuAlaThrCys 187

Search completed: June 18, 2004, 21:04:33
Job time : 24.5 secs
```

Blank